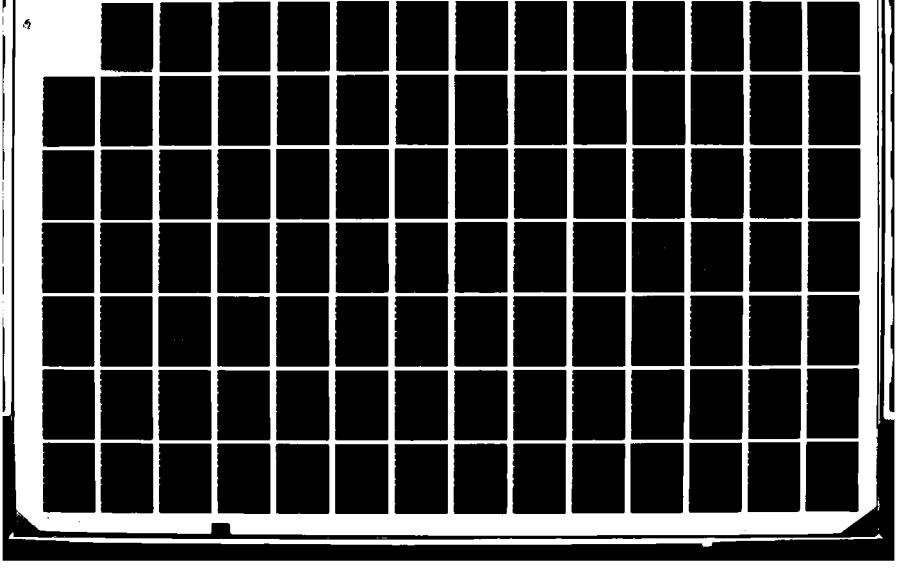
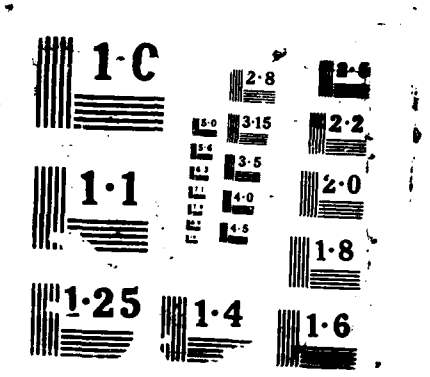
AD-A188 424 PASSIVE ACOUSTIC TRACKING AND FILTERING PROBLEMS WITH TIME SCALES(U) ALPHATECH INC BURLINGTON MA A CAROMICOLI ET AL NOV 87 TR-352 N00014-85-C-0349 1/3

UNCLASSIFIED F/G 17/1 NL




1·C
1·1
1·25
1·4
2·8
3·15
3·5
4·0
4·5
2·2
2·0
1·8
1·6

(12)

TR-352

PASSIVE ACOUSTIC TRACKING AND FILTERING PROBLEMS WITH TIME SCALES

Final Report

ALPHATECH, INC.
2 BURLINGTON EXECUTIVE CENTER
111 MIDDLESEX TURNPIKE
BURLINGTON, MA 01803
617-273-3388

ALPHATECH, INC.

TR-352

PASSIVE ACOUSTIC TRACKING AND FILTERING PROBLEMS WITH TIME SCALES

Final Report

By

Adam Caromicoli
Robert B. Washburn
Alan S. Willsky

November 1987

Supported by

Office of Naval Research

under Contract No. N00014-85-C-0349

ALPHATECH, Inc.
2 Burlington Executive Center
111 Middlesex Turnpike
Burlington, MA 01803
(617) 273-3388

CONTENTS

		Page
FIGURES		iv
1.	INTRODUCTION	1
1.1	PASSIVE ACOUSTIC AIRBORNE ASW	1
1.2	CONVENTIONAL VERSUS INTEGRATED SYSTEM ARCHITECTURE	2
1.3	FILTERING PROBLEMS WITH TIME SCALES	6
1.4	SIGNIFICANCE OF PERTURBATION METHOD	7
1.4.1	Computation Reduction	7
1.4.2	Hierarchies of Architectures	7
1.4.3	Nonintuitive Architectures	8
1.4.4	Reduction to Component Design	8
1.5	OVERVIEW OF REPORT	9
2.	PASSIVE ACOUSTIC TRACKING MODELS	11
2.1	INTRODUCTION	11
2.2	TIME DELAY, DOPPLER EFFECT, AND SOURCE-SENSOR MOTION	11
2.3	ATTENUATION	14
2.4	ASPECT DEPENDENCE	15
2.5	SENSOR DIRECTIVITY	18
2.6	NOISE MODEL	20
2.7	EXAMPLE	20
3.	STOCHASTIC ESTIMATION THEORY	26
3.1	INTRODUCTION	26
3.2	FEATURES OF ACOUSTIC TRACKING MODELS	26
3.3	FILTERING MODELS AND EQUATIONS	29
4.	FILTERING PROBLEMS WITH TIME SCALES	33
4.1	INTRODUCTION	33
4.2	SINGULAR AND NEARLY-SINGULAR FILTERING PROBLEMS	34
4.2.1	Introduction	34
4.2.2	Singular Estimation and Control	34
4.2.3	Singularly-Perturbed and Nearly-Singular Estimation and Control	37

CONTENTS (Continued)

		Page
4.3	FILTERING WITH DYNAMICS POSSESSING NATURAL TIME SCALES	53
4.3.1	Continuous-State Problems	53
4.3.2	Discrete-State Problems	55
5.	FILTER ARCHITECTURE FOR TIME-SCALE APPROXIMATIONS	57
5.1	COMPARISON OF FILTERING TECHNIQUES TO THE REQUIREMENTS OF OUR PROBLEM	57
5.2	FILTER ARCHITECTURES OF EXISTING TECHNIQUES	59
5.2.1	Fast Front End, Slow Back End with Feed-Forward and Feedback	59
5.2.2	Slow Processor with Correction	59
5.2.3	Fast Estimate Followed by Slow	59
5.2.4	Steady-State Solution with Boundary Layer	59
5.2.5	Multiple Lower-Order Filters	60
5.2.6	Fast Analog Processor Followed by Slow Digital Processor	60
6.	FILTERING PROBLEM FORMULATIONS	63
6.1	EXAMPLES SELECTED FOR ANALYSIS	63
6.1.1	Measurements Corrupted by White Noise	63
6.1.2	Perfect Measurements with Small Differences in Rates	65
6.2	CONJECTURED ASYMPTOTIC BEHAVIOR AND FILTER ARCHITECTURES	65
6.2.1	The White Measurement Noise Model	65
6.2.2	Perfect Measurements with Small Differences in Rates	82
7.	ASYMPTOTIC ANALYSIS	92
8.	MONTE CARLO ANALYSIS OF FILTERING PROBLEM	110
8.1	Simulation Techniques	110
8.1.1	Noise Processes	111
8.1.2	Markov Processes	111
8.2	SIMULATION OF DIFFERENTIAL EQUATIONS	113
8.2.1	Simulation in a Noisy Environment	113
8.2.2	Simulation in a Noiseless Environment	120

CONTENTS (Continued)

	Page
8.3 SIMULATION RESULTS	125
8.3.1 Probability Propagation for White Noise Model	125
8.3.2 Quantitative Aspects of White Noise Model Simulations	142
8.3.3 Sample Paths and Probabilities for the Discrete Measurement Case	151
8.3.4 Quantitative Simulations for FE/BE Processor in a Noiseless Environment	169
8.4 MONTE CARLO CHARACTERIZATIONS OF BATCH STATISTICS	181
8.5 DISCUSSION	188
9. CONCLUSIONS	195
REFERENCES	198
APPENDIX A - SOLUTION OF THE WAVE EQUATION WITH MOVING SOURCE	A-1
APPENDIX B - FILTERING EQUATIONS FOR PARTIALLY-OBSERVED FINITE-STATE CONTINUOUS-TIME MARKOV PROCESS	B-1
APPENDIX C - DECISION PROBABILITIES FOR THE HYPOTHESIS TEST OF EQS. 6-57 AND 6-58	C-1

FIGURES

Number		Page
1-1	Generic Passive Sonar Processing System	3
1-2	Optimal, Integrated Processing System Architecture	4
1-3	Suboptimal, Segmented Processing System Architecture	5
1-4	Segmented Processing System Architecture with Feedback	6
2-1	Time Delay Geometry	12
2-2	Aspect Dependence	16
2-3	Sensor Directivity	19
4-1	Structure of Sannuti's "Almost Observable Form"	41
4-2	Distribution of High-Gain Optimal Input Across the Integration Chain	42
4-3	Structure Implied by Estimation Dual of Almost Observable Form	47
4-4	Effect of Low Measurement Noise on Steady-State Error Covariances	51
5-1	Fast Front End, Slow Back End with Feed-Forward and Feedback	61
5-2	Slow Processor with Correction	61
5-3	Fast Estimates Followed by Slow	61
5-4	Steady-State Solution with Boundary Layer	62
5-5	Multiple Lower Order Filters	62
5-6	Fast Analog Processor Followed by Slow Digital Processor (Nonlinear Operations)	62
6-1	A 4-State Markov Process	64
6-2	The 2-State Aggregate Process	70

FIGURES (Continued)

Number		Page
6-3	Architecture of the Optimal Filter.	72
6-4	Architecture of the First Approximate Filter.	72
6-5	A Front-End/Back-End Structure Based on the Approximate Filter in Fig. 6-4.	73
6-6	Front-End/Back-End Structure with Front-End Hypothesis Testing.	74
6-7	Model on Which the Front End of Fig. 6-6 is Based	74
6-8	The Architecture of the Front-End in Fig. 6-6	77
6-9	Front-End/Back-End Structure Arising from Slow Integration of Optimal Estimator.	86
6-10	Hypothesis-Testing Front End with Slow Back-End Estimator . . .	86
6-11	The Left-Right Transition Rates as Functions of x(t).	86
8-1	Exact Probabilities for White Noise Model, $g = 1$, $\varepsilon = .01$, $\Delta t = .1$	126
8-2	Exact Probabilities for White Noise Model, $g = .3$, $\varepsilon = .01$, $\Delta t = .1$.	127
8-3	Exact Probabilities for White Noise Model, $g = .1$, $\varepsilon = .01$, $\Delta t = .1$	128
8-4	Exact Probabilities for White Noise Model, $g = .03$, $\varepsilon = .01$, $\Delta t = .1$	129
8-5	Probabilities for Aggregate White Noise Model, $g = .1$, $\varepsilon = .01$, $t = 0.1$.	130
8-6	Probabilities for Aggregate White Noise Model, $g = .3$, $\varepsilon = 0.01$, $\Delta t = 0.1$.	131
8-7	Probabilities for Aggregate White Noise Model, $g = 1$, $\varepsilon = .01$, $\Delta t = 0.1$.	132
8-8	FE/BE Simulation for White Noise Model with $g = .3$, $\varepsilon = 0.01$, $\Delta t = 1$.	133
8-9	FE/BE Simulation for White Noise Model with $g = .3$, $\varepsilon = .01$, $\Delta t = 3$.	134

FIGURES (Continued)

Number		Page
8-10	FE/BE Simulation for White Noise Model, $g = .3$, $\varepsilon = .01$, $\Delta t = 10$.	135
8-11	FE/BE Simulation for White Noise Model, $g = .3$, $\varepsilon = .01$, $\Delta t = 31$.	136
8-12	FE/BE Simulation for Aggregate and White Noise Model, $g = .3$, $\varepsilon = .01$, $\Delta t = 3$	137
8-13	FE/BE Simulation for Aggregate and White Noise Model, $g = .3$, $\varepsilon = .01$, $\Delta t = 10$.	138
8-14	FE/BE Simulation for Aggregate and White Noise Model, $g = .1$, $\varepsilon = .01$, $\Delta t = 10$.	139
8-15	Plot of Log $\left[\text{M.S.} \left(p^L - \left(\frac{\gamma_2}{\gamma_1+\gamma_2} \right) \right) \right]$ of Exact Filter	144
8-16	Plots of Log (MSE) of Aggregate Filter in White Noise	145
8-17	Plot of Log $\left(\frac{\text{Error}}{\text{Information}} \right)$ for Various Orders of $g(\varepsilon)$.	146
8-18	Plots of $\sup_t \left\| p^L - \frac{\gamma_2}{\gamma_1+\gamma_2} \right\|$ of Exact Filter	148
8-19	Plot of Log [sup(error)] for Approximate Filter in White Noise.	149
8-20	Plot of Log $\left(\frac{\sup \lvert\text{error}\rvert}{\sup \lvert\text{info.}\rvert} \right)$ for Various Orders of $g(\varepsilon)$.	150
8-21	Exact Probabilities for Discrete Model, $g_1(\varepsilon) = 0.03$, $\varepsilon = 0.01$, $\Delta t = 0.1$.	153
8-22	Exact Probabilities for Discrete Model, $g_1(\varepsilon) = 0.1$, $\varepsilon = 0.01$, $\Delta t = 0.1$.	154
8-23	Exact Probabilities for Discrete Model, $g_1(\varepsilon) = 0.3$, $\varepsilon = 0.01$, $\Delta t = 0.1$.	155
8-24	Exact Probabilities for Discrete Model, $g_1(\varepsilon) = 1.0$, $\varepsilon = 0.01$, $\Delta t = 0.1$.	156

FIGURES (Continued)

Number		Page
8-25	Exact Probabilities for Discrete Model, $g_1(\varepsilon) = 3.0$, $\varepsilon = .01$, $\Delta t = 0.1$	157
8-26	Differential Equation Approximation with Time Increment of 1.	158
8-27	Differential Equation Approximation with Time Increment of 10	159
8-28	Differential Equation Approximation with Time Increment of 50	160
8-29	Differential Equation Approximation with Time Increment of 10 and $g_1(\varepsilon) = 0.03$	162
8-30	Differential Equation Approximation with Time Increment of 10 and $g_1(\varepsilon) = 0.3$	163
8-31	FE/BE Simulation for $g_1(\varepsilon) = 0.1$ and Updates Every 1 Second	164
8-32	FE/BE Simulation for $g_1(\varepsilon) = 0.1$ and Updates Every 10 Seconds	165
8-33	FE/BE Simulation for $g_1(\varepsilon) = 0.1$ and Updates Every 31 Seconds	166
8-34	FE/BE Simulation for $g_1(\varepsilon) = 0.1$ and Updates Every 75 Seconds	167
8-35	FE/BE Simulation for $g_1(\varepsilon) = 0.3$ and Updates Every 20 Seconds	168
8-36	Plot of MSE vs. $g(\varepsilon)$ for $T(\varepsilon) = 1$ to 31 and $\varepsilon = 0.1$	170
8-37	Plot of MSE vs. $g(\varepsilon)$ for $T(\varepsilon) = 1$ to 31 and $\varepsilon = .001$	171
8-38	Plot of MSE vs. $T(\varepsilon)$ for $\varepsilon = .001$ to .1 and $g(\varepsilon) = 0.1$	172
8-39	Plot of MSE vs. $T(\varepsilon)$ for $\varepsilon = .001$ to .1 and $g(\varepsilon) = .001$	173
8-40	Plot of MSE vs. $\ell n\left(g(\varepsilon)(1+\varepsilon T^2(\varepsilon))\right)^{1/2}$ for $\varepsilon = .1$	174
8-41	Plot of MSE vs. $\ell n\left(g(\varepsilon)(1+\varepsilon T^2(\varepsilon))\right)^{1/2}$ for $\varepsilon = .01$	175
8-42	Plot of MSE vs. $\ell n\left(g(\varepsilon)(1+\varepsilon T^2(\varepsilon))\right)^{1/2}$ for $\varepsilon = .001$	176
8-43	Plot of MSI vs. ε for $g(\varepsilon)$ from .001 to .1 and $T(\varepsilon) = 1$	177
8-44	Plot of MSI vs. $g(\varepsilon)$ for $\varepsilon = .001$ to .1 and $T(\varepsilon) = 1$.	178
8-45	Plot of MSI vs. $\ell n\left(g^2(\varepsilon)/\varepsilon\right)$ with $T(\varepsilon) = 1$	179
8-46	Plot of $p(\ell \mid L)$, $p(\ell \mid R)$ for $T = 1$, $g_1 = 0.3$ and 200 Runs	182

FIGURES (Continued)

Number		Page
8-47	Plot of $p(\ell\|L)$, $p(\ell\|R)$ for T = 3, g_1 = 0.3 and 200 Runs	183
8-48	Plot of $p(\ell\|L)$, $p(\ell\|R)$ for T = 10, g_1 = 0.3 and 2500 Runs . . .	184
8-49	Plot of $p(\ell\|L)$, $p(\ell\|R)$ for T = 31, g_1 = 0.3 and 2500 Runs . . .	185
8-50	Plot of $p(\ell\|L)$, $p(\ell\|R)$ for T = 100, g_1 = 0.3 and 2500 Runs. . .	186
8-51	Gaussian Approximations for $g(\varepsilon)$ = 0.3, $T(\varepsilon)$ = 1.	189
8-52	Gaussian Approximations for $g(\varepsilon)$ = 0.3, $T(\varepsilon)$ = 3.	190
8-53	Gaussian Approximations for $g(\varepsilon)$ = 0.3, $T(\varepsilon)$ = 10	191
8-54	Gaussian Approximations for $g(\varepsilon)$ = 0.3, $T(\varepsilon)$ = 31	192
8-55	Gaussian Approximations for $g(\varepsilon)$ = 0.3, $T(\varepsilon)$ = 100.	193

SECTION 1

INTRODUCTION

1.1 PASSIVE ACOUSTIC AIRBORNE ASW

The airborne ASW tracking problem is becoming progressively more difficult every year. The continued efforts of our adversaries toward quieter, faster assets with longer ranged weapons imply the need for more powerful localization and tracking systems able to extract maximum information from lower signal strengths and to obtain fire control solutions as quickly as possible. Current systems work by extracting bearing and Doppler information from narrow band signals and then using this information in Extended Kalman filter tracking algorithms. To operate successfully such systems require a stable narrow band line and relatively high signal-to-noise ratio (narrow band signal strength to background noise strength). Unfortunately, these requirements are becoming increasingly more difficult to satisfy as targets become quieter.

Several new sensors planned for the future (e.g., various types of array buoys) will help prolong the usefulness of current tracking systems, but it is clear that within very few years new tracking systems will be required to exploit many different types of measurements and to operate at lower SNR levels than possible with current systems.

The key mathematical problem of passive ASW localization and tracking is the estimation of the state of time-varying stochastic processes given

nonlinear measurements with low SNR. A theoretical framework does exist for such problems, at least at the abstract level, in the current theory of nonlinear estimation. Unfortunately, this theory has very little to say at the practical, computational level about nonlinear problems in general. Nevertheless, useful theoretical results are possible for nonlinear filtering problems if additional mathematical structure is present, and we believe that the passive tracking problem of airborne ASW does have significant special structure to exploit. The research discussed here identifies one type of special structure (time-scale perturbation) and shows how to exploit it to design improved passive tracking systems.

1.2 CONVENTIONAL VERSUS INTEGRATED SYSTEM ARCHITECTURE

Current passive acoustic tracking systems work by extracting target parameters (such as time delay or Doppler shift of a narrow band emission) from a raw signal and then processing those parameters to obtain estimates of target parameters of interest (position and velocity). These two types of processing are referred to as signal processing and tracking. Conventional system designs assume that signal processing and tracking can be performed sequentially (see Fig. 1-1 from [1]). The front-end signal processor is usually designed assuming that target parameters are constant over time; the back-end tracker is designed assuming that the parameter estimates output from the signal processor are direct measurements of the true parameter values with the addition of uncorrelated measurement noise.

These design assumptions are based on the physical nature of observation and target processes: that is, a slowly varying target process modulates a rapidly varying observation process. Thus, the conventional system design

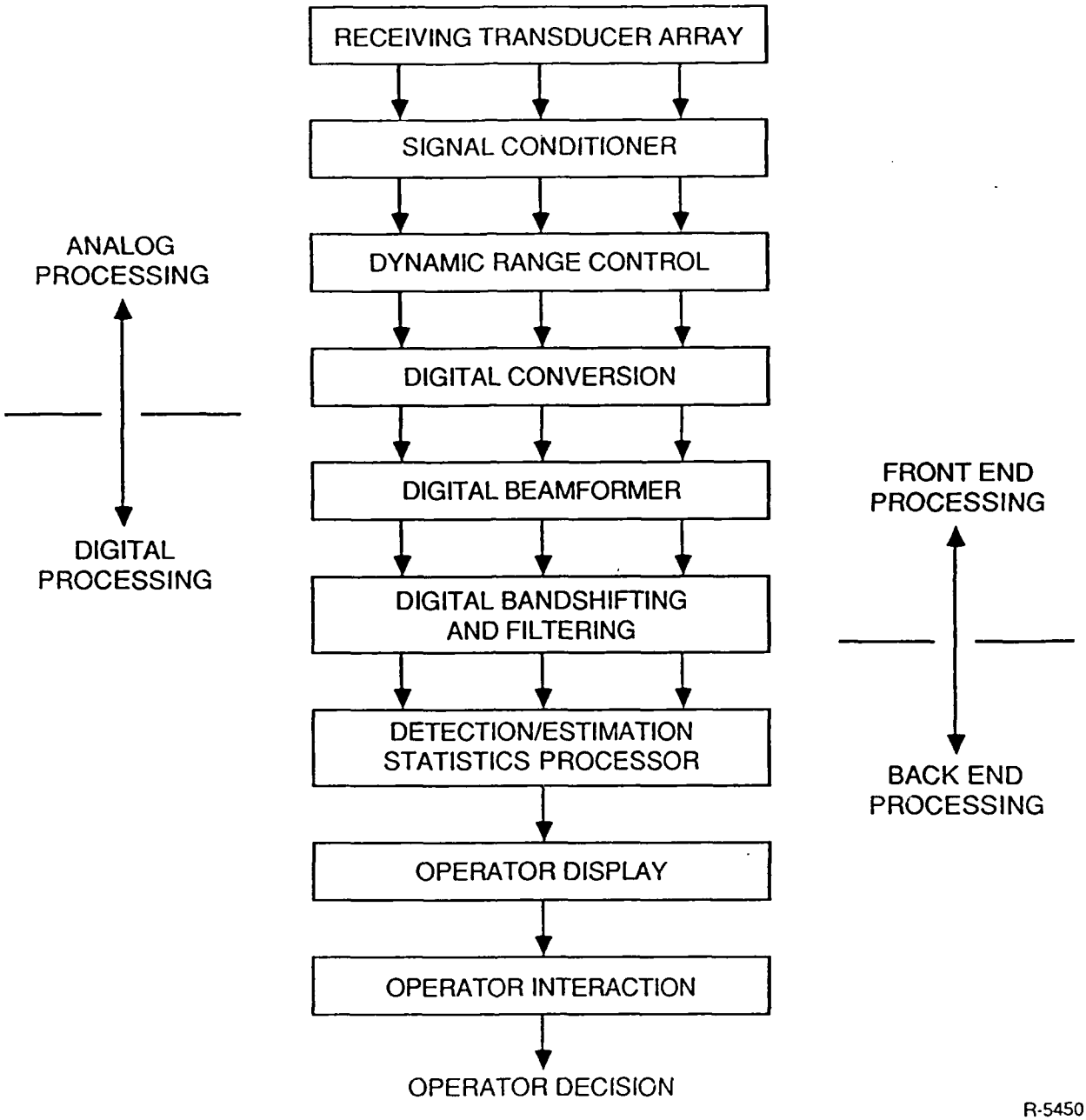


Figure 1-1. Generic Passive Sonar Processing System (Knight, Pridham, Kay, 1981)

inverts the physical model -- i.e., a fast signal processor modulates a slow tracking filter. However, these design assumptions are only approximations and the segmented system is only an approximation of the optimal processor for

extracting target tracks from the raw acoustic sensor data. In an ideal world with unlimited computational and data transmission resources, one would design an optimal, totally integrated tracking system that takes in raw hydrophone signals from every sonobuoy and outputs target position and velocity estimates (Fig. 1-2).

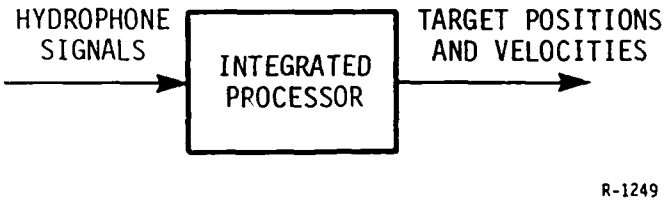


Figure 1-2. Optimal, Integrated Processing System Architecture

We know what the optimal design is -- it is the implementation of Bayes rule. One way to realize this optimal algorithm is to discretize the continuous-valued variables and treat the problem as a finite state Markov estimation problem. As the discretization becomes more refined, this approximation comes closer to the optimal Bayesian algorithm. Unfortunately, the computational complexity also increases so rapidly that it quickly overwhelms any imaginable processor for all but problems of small dimensions (at most 2 or 3). Numerical and processor work on this problem is still advancing [2] but it is clear that this completely optimal, totally integrated approach will have to be used in conjunction with methods that decompose large dimensional problems into a collection of small dimensional problems which can be solved separately and then recombined to obtain a tractable solution to the overall high dimensional problem.

Such decompositions define corresponding system architectures, i.e., a collection of components (which solve the small dimensional subproblems)

together with the inputs and outputs of each component (and thus, connections between components). For example, Fig. 1-3 shows a typical passive tracking system architecture with two components, a front-end signal processor that estimates bearing and frequency lines from raw measurements and a back-end tracking algorithm that estimates position and velocity from bearing and frequency inputs. Figure 1-4 shows another system architecture. It adds a feedback connection from the back-end tracker to the front-end signal processor. This feedback can enhance the signal processing by providing estimates of expected Doppler shift and bearing to the front end. However, if the tracker back end is producing poor position and velocity estimates, this feedback may in fact worsen the performance of the front-end signal processor. How does the system of Fig. 1-4 decide when to switch feedback on or off? More generally, how does one decide which architecture is the best one to use in a particular situation? Are frequency and bearing outputs the best ones to provide to the back-end tracker in Fig. 1-4? If the architecture of Fig. 1-4 is used, what is the best feedback information that the back-end tracker can provide the front-end signal processor in order to tell the front end when it should and when it should not use this information to enhance its signal processing? The objective of this research is to show that we can apply stochastic perturbation theory to answer such questions as these in a precise, systematic manner.

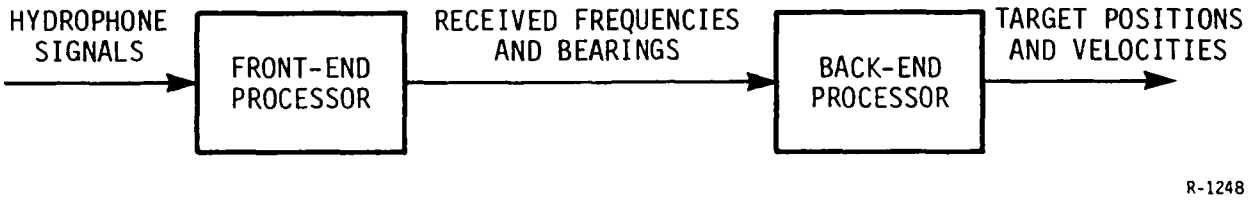


Figure 1-3. Suboptimal, Segmented Processing System Architecture

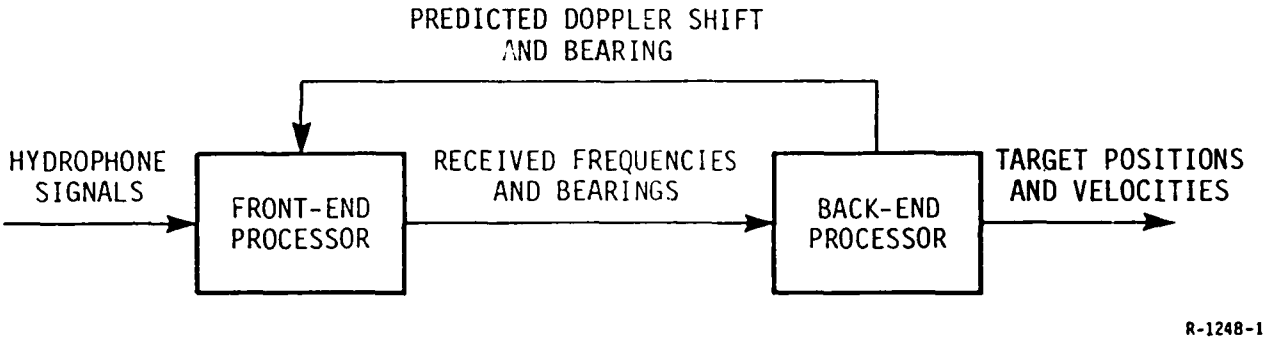


Figure 1-4. Segmented Processing System Architecture with Feedback

1.3 FILTERING PROBLEMS WITH TIME SCALES

Our objective is to employ some systematic mathematical methods to identify architectures (that is, components, inputs, and outputs) which perform close to optimal, totally integrated systems. Our approach to doing this is to formulate the complete tracking problem as a mathematical estimation problem, identify perturbation parameters in the mathematical model, use methods of stochastic perturbation theory to decompose the estimation problem, and interpret the decomposition in terms of a corresponding system architecture.

The specific perturbation parameter we will exploit in this work is the ratio of the signal time constant to the target time constant which is often small. The conventional segmented signal processing and tracking system is an approximation based on this time-scale separation. The goal of our research is to obtain time-scale approximations of optimal filters systematically and rigorously from a mathematical analysis of the equations modeling the target and signal.

1.4 SIGNIFICANCE OF PERTURBATION METHOD

We have mentioned only one kind of perturbation (namely time-scale perturbation) so far, but we believe that several other useful perturbations can be identified in the mathematical models used in passive acoustic tracking. In this subsection we will discuss the significance of developing architecture decomposition methods based on stochastic perturbation theory.

1.4.1 Computation Reduction

As we noted at the beginning of this section, the purpose of this method is to decompose high dimensional problems into tractable lower dimensional problems. What is significant is that perturbation theory provides a systematic, general way of doing this. Moreover, this approach provides decompositions which are small perturbations of the optimal, totally integrated system. How much performance loss results depends on the size of the perturbation parameter. Although quantitative performance analysis is far from easy even in these cases, the method does provide an approach to identifying decompositions which promise minimal loss of performance.

1.4.2 Hierarchies of Architectures

The perturbation method also naturally provides hierarchical families of architectures in terms of asymptotic expansions of optimal solutions. That is, one can identify the simplest architectures resulting from assuming that the perturbation parameters are negligible, but also one can identify more complex architectures which may also have better performance when the perburbation parameter is not negligible. For example, it seems likely that the architecture shown in Fig. 1-3 is a zero-order approximation of the optimal

system when the perturbation parameter is negligible. We also conjecture that some variation of the architecture shown in Fig. 1-4 is a first-order approximation of the optimal system. What we do not know is what the exact feedback information should be in Fig. 1-4, but we expect stochastic perturbation theory to give us the precise answer.

1.4.3 Nonintuitive Architectures

In many problems experienced, intuition suffices to determine zero-order and sometimes high order perturbation approximations. Perturbation theory is much more complicated for stochastic problems, and we believe that a systematic, mathematical method such as we propose will have a tremendous advantage over unaided intuition. Reference [3] gives some examples of nonintuitive stochastic perturbation results.

1.4.4 Reduction to Component Design

The object of architecture *decomposition* is to reduce the problem of designing a tracking system to that of designing simpler components which are connected together as specified by the architecture. Architecture decomposition will specify these components in terms of an estimation subproblem that the component needs to solve. The solution of the subproblem would generally be determined by other methods, and in many cases it may happen that methods already exist for solving these subproblems. Thus, the architecture decomposition shows how to utilize these existing results in a unified tracking system to solve the problem at hand. For example, architecture decomposition would indicate that the narrowband problem above has the architecture of Fig. 1-3. This architecture would specify the frontend only as something

which tries to estimate constant frequency and bearing from hydrophone data. The system designer is then free to note that there already exist various solutions to that problem (e.g., maximum likelihood estimators or maximum entropy methods) and would substitute the appropriate module in his system. If the architecture indicates that this subproblem is particularly difficult (for example, the architecture analysis might indicate that the subproblem involves a low signal-to-noise ratio), then the system designer would know he should search for a more sophisticated front end.

1.5 OVERIVEW OF REPORT

Section 2 of this report describes nonlinear filtering models for passive acoustic tracking which we will use as a basis for our study of time-scale approximations and filter architectures. Section 3 briefly reviews the fundamentals of stochastic estimation for dynamical systems in terms of stochastic differential equations. In Section 4 we review previous work on problems similar to ours in the areas of singular estimation and control, and singularly perturbed estimation and control. Section 5 then contains a discussion of existing techniques for the type of perturbation problem arising in passive acoustic tracking; this section discusses several filter architectures that are implied by the different techniques reviewed in Section 4. One of the conclusions of this section is that there are essentially no existing results for the class of models that includes those arising in passive ASW tracking. The remainder of the report then deals with the development of such results. In Section 6 we introduce two relatively simple nonlinear filtering problems possessing the same qualitative features found in passive ASW problems. We also motivate and define a number of approximate solutions to these problems

and discuss their architectural implications. Section 7 contains the theoretical analysis of one of these approximations that provides a precise relationship between process time scales and measurement quality. The extensive simulations in Section 8 not only support the result of Section 7 but also lead to a number of additional conclusions including several concerning the structure and asymptotic properties of front-end/back-end architectures. The body of the report then concludes with a brief review and discussion in Section 9. Appendix A, B, and C contain technical derivations pertinent to Sections 2, 3, and 6, respectively.

SECTION 2

PASSIVE ACOUSTIC TRACKING MODELS

2.1 INTRODUCTION

This section will describe a model for passive acoustic tracking problems, which has been used for passive tracking algorithms in the past and is useful for applying the nonlinear filtering approach we are taking. The basic simplifying assumption is that acoustic signals travel in straight lines from transmitter to receiver at a constant speed. From this assumption, we develop simple models for source and sensor motion effects, source aspect angle dependence, sensor directivity, and attentuation. To be sure, sound propagation in the ocean is much more complex than the model of it presented here [4]. But including a more realistic model of sound propagation, although possible, is well beyond the scope and needs of this research. The model presented here has been sufficient to guide and test our research into new methods of combined signal processing and tracking which may prove useful for practical passive acoustic tracking.

2.2 TIME DELAY, DOPPLER EFFECT, AND SOURCE-SENSOR MOTION

Define the following notation (see Fig. 2-1):

$x_T(t)$ = Transmitter location at time t;

$x_R(t)$ = Receiver location at time t;

$y_T(t)$ = Signal transmitted at time t;

$y_R(t)$ = Signal received at time t.

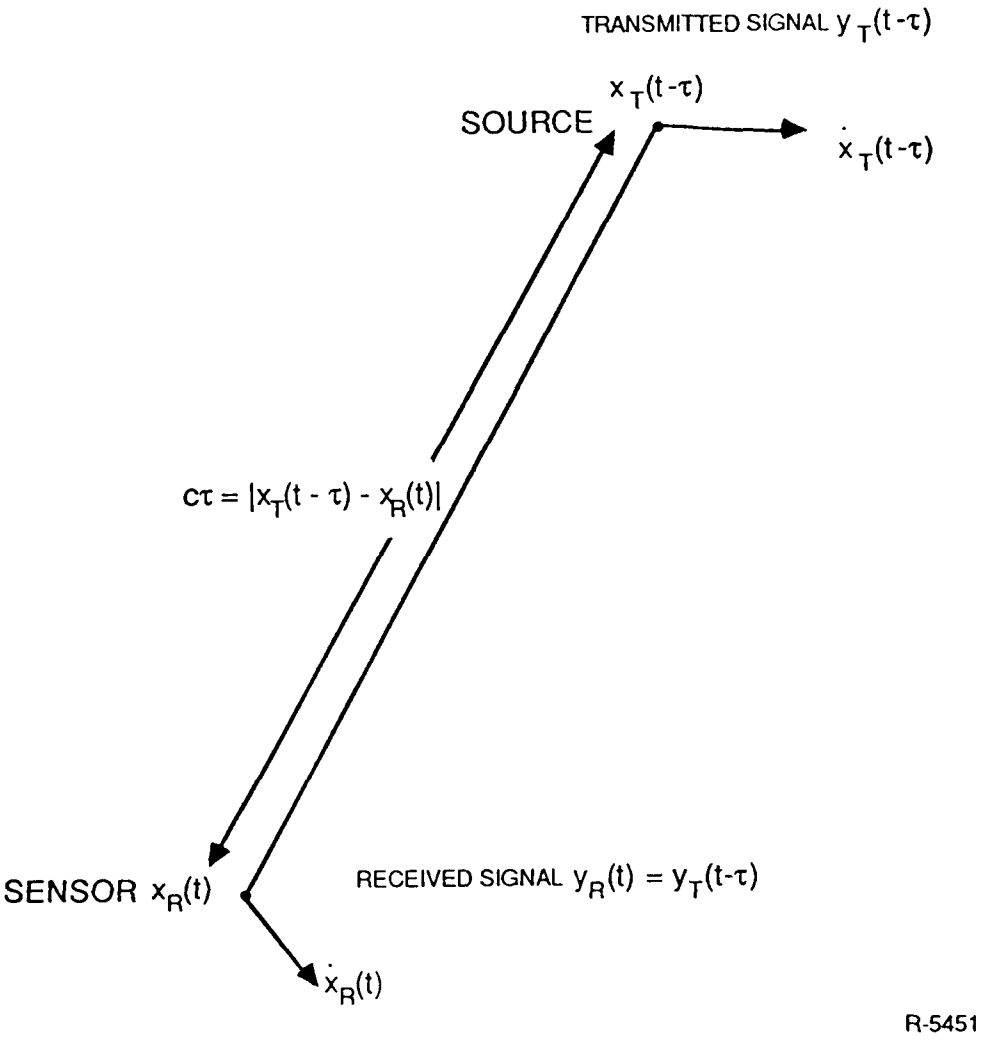


Figure 2-1. Time Delay Geometry

Define $\tau(t)$ implicitly as

$$\begin{aligned} \tau(t) &= \text{Delay in signal } \underline{\text{received}} \text{ at time } t; \\ &= \frac{1}{c} \left| x_T(t-\tau(t)) - x_R(t) \right| \quad . \end{aligned} \tag{2-1}$$

The basic relationship between y_T and y_R in this case is

$$y_R(t) = y_T(t-\tau(t)) \quad . \tag{2-2}$$

The time derivative $\dot{\tau}$ of the time delay τ and the <u>Doppler effect factor</u> $1-\dot{\tau}$ are of particular interest. These are given by the following equations obtained by implicit differentation of Eq. 2-1:

$$\dot{\tau} = \frac{\langle x_T(t-\tau) - x_R(t), \dot{x}_T(t-\tau) - \dot{x}_R(t) \rangle}{c \left| x_T(t-\tau) - x_R(t) \right| \cdot \left[1 + \frac{\langle x_T(t-\tau) - x_R(t), \dot{x}_T(t-\tau) \rangle}{c \left| x_T(t-\tau) - x_R(t) \right|} \right]} \tag{2-3}$$

where $\langle \ , \ \rangle$ denotes the inner product of vectors, and

$$1-\dot{\tau} = \frac{\left[1 + \frac{\langle x_T(t-\tau) - x_R(t), \dot{x}_R(t-\tau) \rangle}{c \left| x_T(t-\tau) - x_R(t) \right|} \right]}{\left[1 + \frac{\langle x_T(t-\tau) - x_R(t), \dot{x}_T(t-\tau) \rangle}{c \left| x_T(t-\tau) - X_R(t) \right|} \right]} \tag{2-4}$$

If the sensor is motionless, then Eqs. 2-3 and 2-4 reduce to

$$\dot{\tau} = \frac{\langle x_T(t-\tau) - x_R, \dot{x}_T(t-\tau) \rangle}{c \left| x_T(t-\tau) - x_R \right| \cdot \left[1 + \frac{\langle x_T(t-\tau) - x_R, \dot{x}_T(t-\tau) \rangle}{c \left| x_T(t-\tau) - x_R \right|} \right]} \tag{2-5}$$

$$1-\dot{\tau} = \frac{1}{\left[1 + \frac{\langle x_T(t-\tau) - x_R, \dot{x}_T(t-\tau)\rangle}{c\ |x_T(t-\tau) - x_R|}\right]} \quad . \tag{2-6}$$

If the target is motionless, then Eqs. 2-3 and 2-4 reduce to

$$\dot{\tau} = \frac{-\langle x_T - x_R(t), \dot{x}_R(t)\rangle}{c\ |x_T - x_R(t)|} \quad , \tag{2-7}$$

$$1-\dot{\tau} = 1 + \frac{\langle x_T - x_R(t), \dot{x}_R(t)\rangle}{c\ |x_T - x_R(t)|} \quad . \tag{2-8}$$

The time delay model presented here in Eq. 2-2 is an approximation to the exact solution of the three-dimensional wave equation for a moving point source. The exact solution for the case of a stationary receiver ($\dot{x}_R = 0$) is derived in Appendix A and is given by

$$y_R(t) = y_T(t-\tau) \cdot \frac{(1-\dot{\tau})}{c\tau} \quad . \tag{2-9}$$

where $y_T(t)$ is the sound pressure level at the point source and $y_R(t)$ is the sound pressure level at the receiver.

2.3 ATTENUATION

Real signals are attenuated as they propagate. Such effects can be added in the time delay model by defining an attenuation function

$A(x_T, x_R)$ = ratio of transmitted to received signal amplitude for transmitter at x_T and receiver at x_R.

The corresponding received signal model is

$$y_R(t) = A\left[x_T(t-\tau), x_R(t)\right]\ y_T(t-\tau) \quad . \tag{2-10}$$

The attenuation function might satisfy a power law

$$A(x_T, x_R) = |x_T - x_R|^{\gamma} \tag{2-11}$$

where $\gamma = -2$ for spherical spreading (propagation in three dimensions), $\gamma = -1$ for cylindrical spreading (propagation in two dimensions), or something in between ($-2 < \gamma < -1$). Urick [4] notes some values of γ to use in the sonar propagation loss function (which is essentially what is being modeled).

Note that the exact solution (Eq. 2-9) for the homogeneous three-dimensional case includes the geometric attenuation due to spherical spreading (the factor appears only as $(c\tau)^{-1} = r^{-1}$ attenuation of sound pressure level).

2.4 ASPECT DEPENDENCE

Figure 2-2 shows the geometry of aspect dependence of the transmitted signal y_T. Note that in this section we assume that receiver and transmitter are moving in a plane. The notation is

$x_T(t)$ = Transmitter location at time t;

$x_R(t)$ = Receiver location at time t;

$y_T(t)$ = Signal transmitted at time t;

$y_R(t)$ = Signal received at time t;

$\phi_a(t)$ = Acoustic (delayed) aspect angle of target at time t;

$\tau(t)$ = Delay in signal received at time t;

$g(\phi)$ = Directivity of transmitted signal relative to target heading.

The transmitted and received signals are related by

$$y_R(t) = g\big(\phi_a(t)\big)\, y_T(t-\tau) \quad . \tag{2-12}$$

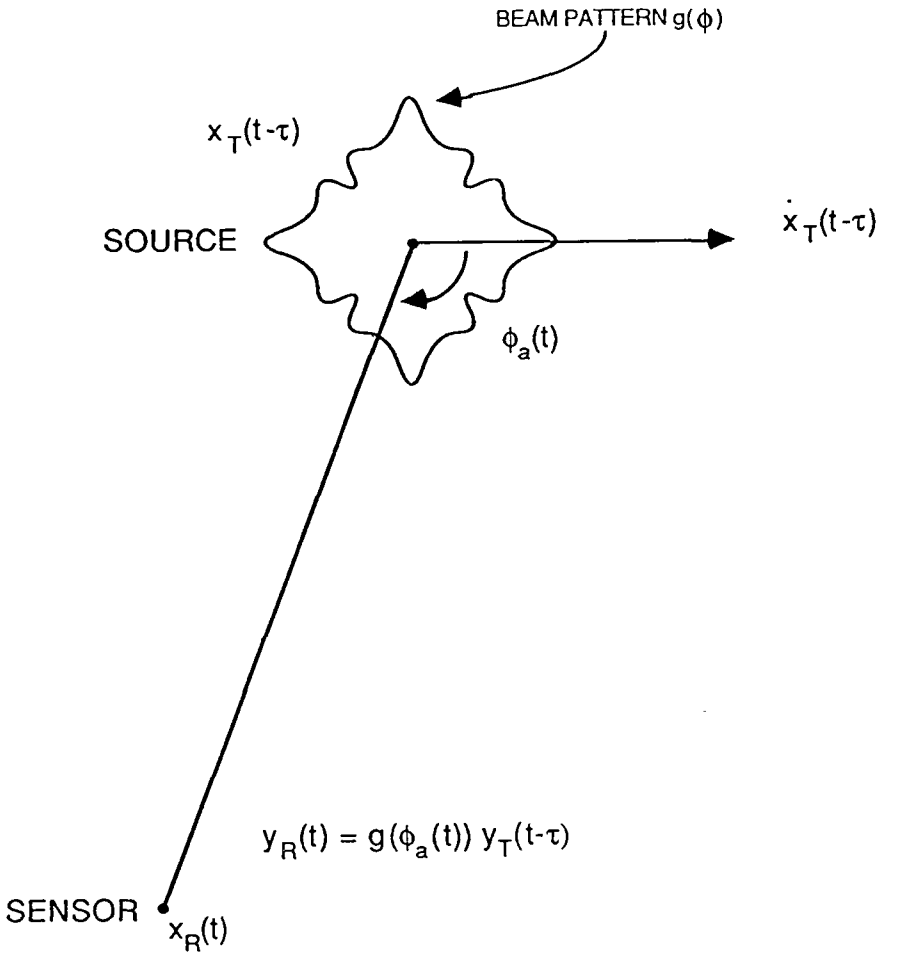


Figure 2-2. Aspect Dependence

Note that ϕ_a satisfies

$$\cos\left(\phi_a(t)\right) = \frac{-\langle x_T(t-\tau) - x_R(t), \dot{x}_T(t-\tau)\rangle}{|x_T(t-\tau) - x_R(t)|\ |\dot{x}_T(t-\tau)|} \quad . \tag{2-13}$$

If the source transmits several different signals $y_{T,k}$ with different directivity functions g_k, then the received signal is the superposition of signals (Eq. 2-12). That is,

$$y_R(t) = \sum_{k=1}^{n} g_k\left(\phi_a(t)\right)\, y_{T,k}(t-\tau) \quad . \tag{2-14}$$

To treat transmitters and receivers moving in three dimensions, define

$b_d(t)$ = delayed bearing directional vector at time t,

$$= \frac{x_T(t-\tau) - x_R(t)}{|x_T(t-\tau) - x_R(t)|} \quad ;$$

$v_T(t)$ = target heading directional vector at time t,

$$= \frac{\dot{x}_T(t)}{|\dot{x}_T(t)|} \quad .$$

The source directivity g should be a function of b and v_T so that the received signal is given by

$$y_R(t) = g\left(b_d(t), v_T(t-\tau)\right)\, y_T(t-\tau) \quad . \tag{2-15}$$

If the receiver is motionless, then the delayed bearing is related to the instantaneous bearing by the equation

$$b_d(t) = b(t-\tau) \tag{2-16}$$

where b(t) is defined as

$b(t)$ = instantaneous directional bearing vector,

$$= \frac{x_T(t) - x_R}{|x_T(t) - x_R|} \quad .$$

2.5 SENSOR DIRECTIVITY

Sensor directivity can be treated similar to aspect angle dependence. Figure 2-3 shows the geometry; as in the previous section, we assume that receiver and transmitter are coplanar. Define the notation

$x_T(t)$ = Transmitter location at time t;

$x_R(t)$ = Receiver location at time t;

$y_T(t)$ = Signal transmitted at time t;

$y_R(t)$ = Signal received at time t;

$\beta_a(t)$ = Acoustic (delayed) bearing angle to target at time t;

$\tau(t)$ = Delay in signal received at time t;

$h(\phi)$ = Directivity of received signal relative to receiver velocity or fixed reference direction.

The relation between transmitted and received signals is similar to Eq. 2-12:

$$y_R(t) = h\big(\beta_a(t)\big)\, y_T(t-\tau) \quad . \tag{2-17}$$

Note that β_a satisfies

$$\cos\big(\beta_a(t)\big) = \frac{\langle x_T(t-\tau) - x_R(t), \dot{x}_R(t)\rangle}{|x_T(t-\tau) - x_R(t)|\ |\dot{x}_R(t)|} \quad . \tag{2-18}$$

If the receiver is not moving, replace $\dot{x}_R$ by a reference vector in Eq. 2-18.

A source transmitting multiple signals is modeled by

$$y_R(t) = h\big(\beta_a(t)\big) \sum_{k=1}^{n} y_{T,k}(t-\tau) \quad . \tag{2-19}$$

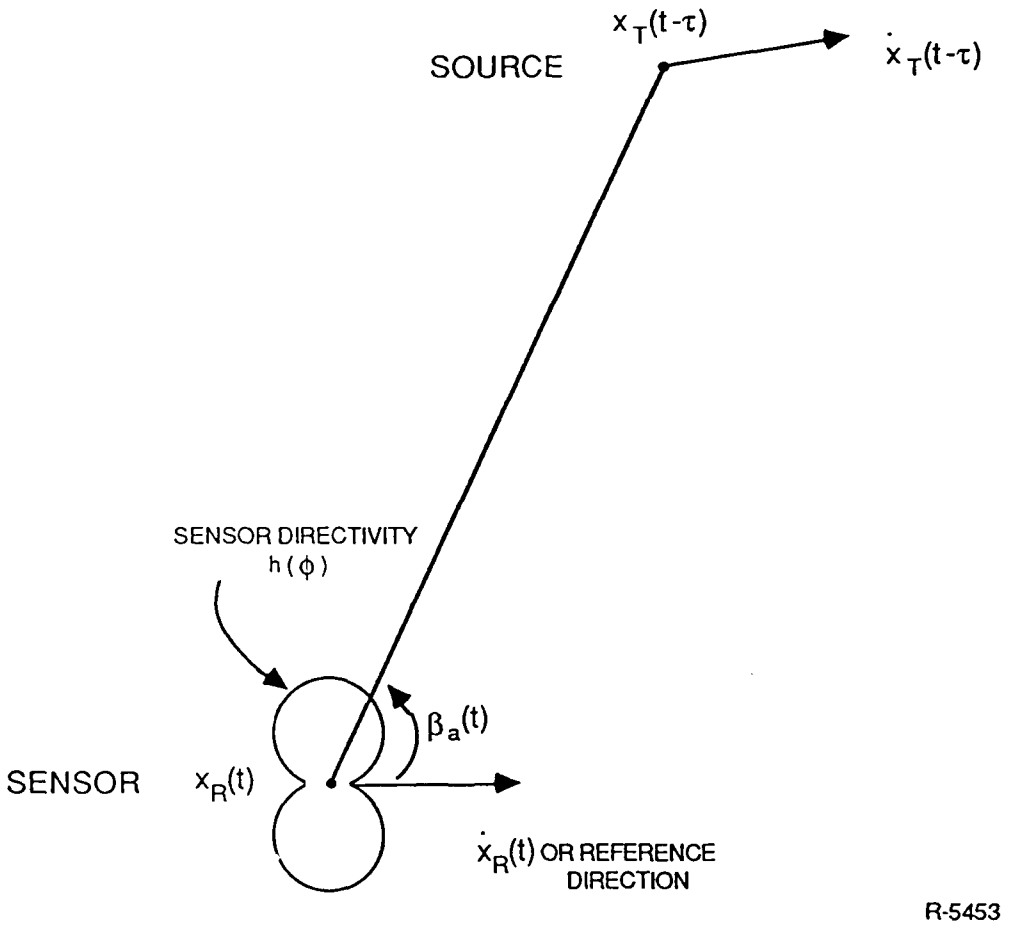


Figure 2-3. Sensor Directivity

2.6 NOISE MODEL

In this section we will only present a model for ambient background noise, which we model as an additive noise term in the equation for the received signal. Define

$n(t)$ = Ambient background noise near the receiver at time t;

$y_T(t)$ = Signal transmitted at time t;

$y_R(t)$ = Signal received at time t;

$\tau(t)$ = Delay in signal received at time t; relative to receiver velocity or fixed reference direction.

The relationship between the received signal and the transmitted signal and background noise is given by

$$y_R(t) = y_T(t-\tau) + n(t) \quad . \tag{2-20}$$

Note that the noise process n(t) in Eq. 2-20 is not delayed since it does not originate from the target. Note also that nothing is said about how n(t) depends on the receiver location. One can include a model of the correlation such as

$$E\{n_1(t)n_2(s)\} = R(x_1(t)-x_2(s),t-s) \tag{2-21}$$

for two receivers of the same type located at x_1 and x_2.

2.7 EXAMPLE

To illustrate the time delay model described above, consider a single source moving with constant velocity and emitting a narrowband signal. The source trajectory satisfies the differential equation.

$$\frac{d^2x_T}{dt^2} = 0 \quad . \tag{2-22}$$

The source signal satisfies the stochastic differential equation

$$d\phi_T = f\ dt + \sigma\ dw \tag{2-23}$$

where f is the constant source frequency, w is a standard Wiener process, and the transmitted signal is

$$y_T(t) = \sin\ \phi_T(t) \quad . \tag{2-24}$$

The signal y_T is a (wide-sense) stationary process with total power 1/2 and two-sided power spectral density at frequency ω given by

$$\frac{\sigma^2}{4\pi} \frac{\frac{\sigma^4}{4} + f^2 + \omega^2}{\left[\frac{\sigma^4}{4} + (f-\omega)^2\right]\left[\frac{\sigma^4}{4} + (f+\omega)^2\right]} \quad . \tag{2-25}$$

This expression represents a spectrum with peak at frequency f (or f and -f for the two-sided spectrum) and width proportional to σ^2.

Now suppose that there is a motionless sensor located at x_R. Consider a "difar" -- i.e., a sonobuoy with (limited) horizontal directivity. Here we assume a two-dimensional model and the difar measures three things:

an omnidirectional measurement,

$$dy_{R,1} = A_1\ y_T(t-\tau)\ dt + dn_1 \quad ; \tag{2-26}$$

and two direction measurements,

$$dy_{R,2} = A_2\ \cos\big(\beta(t-\tau)\big)\ y_T(t-\tau)\ dt + dn_2 \tag{2-27}$$

$$dy_{R,3} = A_3 \sin(\beta(t-\tau))\, y_T(t-\tau)\, dt + dn_3 \quad . \tag{2-28}$$

In these equations, τ is the time delay obtained from

$$\tau = \frac{1}{c} |x_T(t-\tau) - x_R| \tag{2-29}$$

The bearing angle β is defined by the equation

$$\begin{bmatrix} \cos(\beta(t)) \\ \sin(\beta(t)) \end{bmatrix} = \frac{x_T(t)-x_R}{|x_T(t)-x_R|} \tag{2-30}$$

We assume that n_1, n_2, n_3 are standard Wiener processes so that the parameters $(A_1)^2$, $(A_2)^2$, $(A_3)^2$ are signal-to-noise ratios (signal power divided by noise power, where power equals mean square average). Because x_T is a constant velocity trajectory, we can use

$$x_T(t) = x_T(0) + t\, \dot{x}_T(0) \tag{2-31}$$

to solve explicitly for τ as a function of t, $x_T(0)$, and $\dot{x}_T(0)$. If we use the notation

$$v = \dot{x}_T(0) \tag{2-32}$$

$$u = x_T(0)-x_R + t\, \dot{x}_T(0) \quad , \tag{2-33}$$

then τ is given the formula

$$\tau = \frac{\frac{1}{c^2}\langle v,u\rangle + \left[\frac{1}{c^2}\langle v,u\rangle^2 + \left(1 - \frac{|v|^2}{c^2}\right)\cdot\frac{|u|^2}{c^2}\right]^{1/2}}{1 - \left(\frac{|v|^2}{c^2}\right)} \tag{2-34}$$

where $\langle v,u\rangle$ is the scalar product between vectors u and v.

We will now show how to express this model in terms of stochastic differential equations without explicit time delays such as $t-\tau$. Let $\phi_R(t)$ denote $\phi_T(t-\tau)$. Then we can rewrite Eqs. 2-26, 2-27, and 2-28 as

$$dy_{R,1} = A_1 \sin\left(\phi_R(t)\right) dt + dn_1 \tag{2-35}$$

$$dy_{R,2} = A_2 \cos\left(\beta(t-\tau)\right) \sin\left(\phi_R(t)\right) dt + dn_2 \tag{2-36}$$

$$dy_{R,3} = A_3 \sin\left(\beta(t-\tau)\right) \sin\left(\phi_R(t)\right) dt + dn_3 \quad . \tag{2-37}$$

According to McKean [5, p. 41], $\phi_R(t)$ satisfies the stochastic differential equation

$$d\phi_R = f(1-\dot{\tau})dt + \sigma\,(1-\dot{\tau})^{1/2}\, d\tilde{w} \tag{2-38}$$

where $\tilde{w}$ is a standard Wiener process. In other words, the Doppler effect not only shifts the frequency f but it also expands the bandwidth of the signal. The factor $(1-\dot{\tau})^{1/2}$ may not matter much if $\dot{\tau}$ is very small. However, for broadband source signals, this factor is very critical in allowing one to estimate any Doppler effects from the received signal.

To summarize, the time delay model gives us the following differential equation model for a single target and a single directional sensor:

$$\frac{d^2 x_T}{dt^2} = 0 \tag{2-39}$$

$$1-\dot{\tau} = \frac{1}{\left[1 + \dfrac{\langle x_T(t-\tau) - x_R, \dot{x}_T(t-\tau)\rangle}{c\ |x_T(t-\tau)-x_R|}\right]} \tag{2-40}$$

$$d\phi_R = f(1-\dot{\tau})dt + \sigma\,(1-\dot{\tau})^{1/2}\,d\tilde{w} \tag{2-41}$$

$$dy_{R,1} = A_1 \sin\big(\phi_R(t)\big)\,dt + dn_1 \tag{2-42}$$

$$dy_{R,2} = A_2 \cos\big(\beta(t-\tau)\big)\sin\big(\phi_R(t)\big)\,dt + dn_2 \tag{2-43}$$

$$dy_{R,3} = A_3 \sin\big(\beta(t-\tau)\big)\sin\big(\phi_R(t)\big)\,dt + dn_3 \tag{2-44}$$

where τ is given by

$$\tau = \frac{1}{c}\,|x_T(t-\tau) - x_R| \quad , \tag{2-45}$$

and β is given by

$$\begin{bmatrix}\cos\big(\beta(t)\big)\\ \sin\big(\beta(t)\big)\end{bmatrix} = \frac{x_T(t)-x_R}{|x_T(t)-x_R|} \quad . \tag{2-46}$$

In [6],[7] we used a similar model that approximated $\tau = 0$ in Eqs. 2-40, 2-43, 2-44, and 2-45, and ignored the $(1-\dot{\tau})^{1/2}$ factor in Eq. 2-41. This gives the approximate model

$$\frac{d^2x_T}{dt^2} = 0 \tag{2-47}$$

$$1-\dot{\tau} = \frac{1}{\left[1 + \dfrac{\langle x_T(t) - x_R, \dot{x}_T(t)\rangle}{c\,|x_T(t) - x_R|}\right]} \tag{2-48}$$

$$d\phi_R = f(1-\dot{\tau})dt + \sigma\,d\tilde{w} \tag{2-49}$$

$$dy_{R,1} = A_1 \sin\big(\phi_R(t)\big)\,dt + dn_1 \tag{2-50}$$

$$dy_{R,2} = A_2 \cos(\beta(t)) \sin(\phi_R(t)) \, dt + dn_2 \tag{2-51}$$

$$dy_{R,3} = A_3 \sin(\beta(t)) \sin(\phi_R(t)) \, dt + dn_3 \tag{2-52}$$

We will use this example to motivate our study of filter architectures and time-scale approximations in subsequent sections of this report. In the next section we will describe perturbation parameters that can be introduced into this model and which we will exploit to obtain reduced order approximations of optimal filters for problems such as this.

SECTION 3

STOCHASTIC ESTIMATION THEORY

3.1 INTRODUCTION

The first part of this section describes some special features of the passive acoustic tracking model that we can exploit to design approximate optimal filters. These features are not limited to acoustic tracking problems and they are present in many other signal processing and estimation problems. For this reason we will formulate a general class of stochastic perturbation problems which will form the basis of our research. Our work relies on the methods of stochastic differential equations and nonlinear filtering theory [8]. The second part of this section reviews the basic models and results of this theory.

3.2 FEATURES OF ACOUSTIC TRACKING MODELS

The example model described in subsection 2.7 has a general quasilinear form described by three coupled stochastic differential equations:

$$dx = Fx\ dt + G\ dw \tag{3-1}$$

$$dz = A(x)z\ dt + B(x)\ du \tag{3-2}$$

$$dy = C(x)z\ dt + D(x)\ dv \tag{3-3}$$

In these equations w, u, v are assumed to be independent Brownian motion processes. The first equation (Eq. 3-1) describes the target dynamics and

the second two equations (Eqs. 3-2 and 3-3) describe the dynamics of the observed signal y(t) and its dependence on the target state x(t). The matrix coefficients A, B, C, and D depend on the target state x in order to account for such effects as Doppler shifted frequency, directivity of the sensor, and aspect dependence of the target emissions.

Signal processing and tracking problems often have a signal process y(t) with a shorter time constant than the target process x(t). This is the case in passive acoustic tracking where the target dynamics are very slow relative to the acoustic signal dynamics. We can make this time scale explicit by perturbing the stochastic differential equations as follows.

$$dx = Fx\ dt + G\ dw \tag{3-4}$$

$$\varepsilon \cdot dz = A(x)z\ dt + \varepsilon^{1/2} \cdot B(x)\ du \tag{3-5}$$

$$\varepsilon \cdot dy = C(x)z\ dt + \varepsilon^{1/2} \cdot D(x)\ dv \tag{3-6}$$

The parameter ε is proportional to the time constant of the signal process. Note that the square root $\varepsilon^{1/2}$ arises because we are using white noise (du and dv) to drive these stochastic differential equations; speeding up a Brownian motion is mathematically equivalent to scaling its magnitude.

The equations in Eqs. 3-5 and 3-6 are singularly perturbed. That is, their small $\varepsilon \neq 0$ behavior cannot be approximated by simply setting $\varepsilon = 0$. Nevertheless, there are ways to obtain good approximations to the solution of singularly-perturbed problems for small ε. There is a vast literature on singularly-perturbed differential equations. The papers [9],[10] review the work relevant to control and estimation applications. We will review work specifically relevant to our estimation problem in the next section.

As ε approaches 0 in Eqs. 3-5 and 3-6, the rate of information increases to infinity. For example, suppose there is no z process and we have only the two simpler equations

$$dx = Fx\ dt + G\ dw \quad , \tag{3-7}$$

$$\varepsilon \cdot dy = C(x)\ dt + \varepsilon^{1/2} \cdot D(x)\ dv \quad , \tag{3-8}$$

where y is measured and we want to estimate x. This estimation problem is equivalent to one with equations

$$dx = Fx\ dt + G\ dw \quad , \tag{3-9}$$

$$dy = C(x)\ dt + \varepsilon^{1/2} \cdot D(x)\ dv \quad . \tag{3-10}$$

This is an example of an <u>almost singular filtering</u> problem, which we will discuss in Section 4. Note that the signal-to-noise ratio (Eq. 3-10) is proportional to ε^{-1}. Thus, as the signal process becomes faster relative to the target process, the SNR of the signal becomes larger.

Passive acoustic signals typically have low SNR, and a model which predicts increasing SNR is not realistic. Thus, it is necessary to modify the perturbation above to control the SNR. Note that the noise coefficients B and D in Eqs. 3-5 and 3-6 are likely to be large and we cannot let the time scale ε be arbitrarily small independent of these noise factors. Thus, we modify Eqs. 3-5 and 3-6 by adding two functions, $g_1(\varepsilon)$ and $g_2(\varepsilon)$ to allow different scalings of the noise.

$$\varepsilon \cdot dz = A(x)z\ dt + g_1(\varepsilon) \cdot B(x)\ du \tag{3-11}$$

$$\varepsilon \cdot dy = C(x)z\ dt + g_2(\varepsilon) \cdot d(x)\ dv \tag{3-12}$$

If $g_1(\varepsilon) = g_2(\varepsilon) = \varepsilon^{1/2}$, we have the pure time-scale perturbation. We can account for greater noise by allowing different functions of ε. For example, $g_2(\varepsilon) = 1$ causes the SNR to remain constant as $\varepsilon \to 0$, and $g_2(\varepsilon) = \varepsilon^{-1/2}$ causes SNR to decrease to 0 as $\varepsilon \to 0$. In Section 4 we will survey previous work on perturbed filtering problems with an eye to different possible choices for g_1, g_2. Before doing that, let us first review the basic results of nonlinear filtering theory which is the basis of our approach.

3.3 FILTERING MODELS AND EQUATIONS

The model of passive acoustic tracking described above is an example of a general class of nonlinear, continuous state, continuous time processes driven by white noise. This class has the generic form

$$dx = f(x)\,dt + g(x)\,dw \tag{3-13}$$

$$dy = h(x)\,dt + dv \tag{3-14}$$

where w and v are independent multidimensional Brownian motion processes. The filtering problem is to estimate x(t) given the measurements y(s) for $0 \leq s \leq t$. The filtering equations in the general nonlinear case are expressed are expressed as follows. Let ϕ be a function of the state variable x. Define $\pi_t(\phi)$ to be the conditional expectation of $\phi(x(t))$ given the measurements up to time t. The filtering equation expresses $\pi_t(\phi)$ in terms of a stochastic differential equations as follows [11],[12]:

$$d\pi_t(\phi) = \pi_t(L\phi) + [\pi_t(h\phi) - \pi_t(h)\pi_t(\phi)]\,d\nu_t \tag{3-15}$$

where ν_t is the innovations process satisfying the stochastic differential equation

$$dv_t = dy(t) - \pi_t(h)dt \tag{3-16}$$

and L is the infinitesimal generator of the process x(t) [11]. In all but a few cases, the solution of Eq. 3-15 requires an infinite dimensional calculation. Nevertheless, there are several types of additional models which have finite dimensional solutions and that can be used to help gain insight into the essential behavior of the filtering problem. In this subsection we will describe these models and the related filtering equations that apply to them.

The simplest case is the familiar linear Gaussian model which has equations

$$dx = Ax\,dt + B\,dw \tag{3-17}$$

$$dy = Cx\,dt + D\,dv \quad . \tag{3-18}$$

In this case w and v are assumed to be independent Brownian motion processes. The optimal filter in this case is the well-known Kalman-Bucy filter [8] which expresses the conditional probability density of x in terms of its mean $\hat{x}$ and covariance P. These satisfy the equations

$$d\hat{x} = A\hat{x}\,dt + P(DRD^T)^{-1}C^T(dy - C\hat{x}dt) \quad , \tag{3-19}$$

$$dP = (AP + PA^T + BQB^T - PC^T(DRD^T)^{-1}CP)\,dt \quad . \tag{3-20}$$

Note that Eq. 3-19 is a stochastic differential equation but that Eq. 3-20 is purely deterministic.

Linear Gaussian models are useful because they can often be used in practice to solve nonlinear problems by linearization. Note that the acoustic tracking model is quasilinear as indicated in Eqs. 3-1, 3-2, and 3-3, and

linearization techniques may be helpful in this case. Nevertheless, linear models and methods are limited and are known to perform poorly in nonlinear problems with low SNR. Furthermore, methods for analyzing linear problems center around the analysis of the Riccati equation (Eq. 3-20) and do not generalize to nonlinear problems. For this reason we have also chosen two types of discrete state problems to analyze. In these problems x(t) is a finite-state continuous-time Markov process and the filtering equations are a finite dimensional system of stochastic differential equations.

A finite-state Markov process is described in terms of its transition rate matrtix Λ, where the elements $\lambda(\xi_2|\xi_1)$ of Λ are the transition rates

$$\Pr\{x(t+dt)=\xi_2|x(t)=\xi_1\} = \delta_{\xi_1\xi_2} + \lambda(\xi_2|\xi_1)\cdot dt + o(dt). \tag{3-21}$$

If p_t denotes the vector of probabilities $\Pr\{x(t)=\xi\}$, then p_t satisfies the linear differential equation

$$dp_t = \Lambda p_t\, dt \quad . \tag{3-22}$$

In the first discrete model we use the same measurement equations as Eq. 3-14, and it is discussed in [11]-[13]. The filtering equation (Eq. 3-15) still applies, but it is finite-dimensional in this case. Let π_t denote the vector of conditional probabilities $\Pr\{x(t)=\xi|y(s),0\leq s\leq t\}$. Then π_t satisfies the stochastic differential equation

$$d\pi_t = \Lambda\pi_t + [h^*\pi_t - \pi_t{}^T h\pi_t]\, d\nu_t \tag{3-23}$$

where h is the vector with components $h(\xi)$, h^* is the diagnonal matrix with elements $h(\xi)$, and ν_t satisfies

$$d\nu_t = dy(t) - h^T \pi_t \, dt \quad . \tag{3-24}$$

The second type of discrete model we will investigate has discrete measurements as well as a discrete state. In this model y(t) is given by

$$y(t) = h(x(t)) \quad , \tag{3-25}$$

and we call this a partially observed Markov process. In this case the filtering equation is a finite-dimensional stochastic differential equation driven by a jump process J_t (which is the number of jumps of y(s) in the interval [0,t]):

$$d\pi_t(\xi) = F_t \, dt + G_t \, dJ_t \tag{3-26}$$

where F_t and G_t are given by

$$F_t = \delta_{h(\xi)y(t)} \cdot \left[\sum_{\xi'} \lambda(\xi|\xi')\pi_t(\xi') - \pi_t(\xi) \sum_{\xi'} \lambda(y(t)|\xi')\pi_t(\xi') \right] \tag{3-27}$$

where

$$\lambda(y(t)|\xi) = \sum_{h(\xi')=y(t)} \lambda(\xi'|\xi) \quad ,$$

$$G_t = \delta_{h(\xi)y(t)} \cdot \frac{\sum_{\xi'} \lambda(\xi|\xi')\pi_{t-}(\xi')}{\sum_{\xi'} \lambda(y(t)|\xi')\pi_{t-}(\xi')} - \pi_{t-}(\xi) \quad . \tag{3-28}$$

Appendix B derives this filtering equation and discusses several variations of the model and of the filtering equation that will be used in studying this type of problem.

SECTION 4

FILTERING PROBLEMS WITH TIME SCALES

4.1 INTRODUCTION

In this section we review previous work on stochastic filtering problems with multiple time scales that has some relevance to our problem. Although no previous work addresses the problem of interest to us, some of this work is related and the methods used serve as a point of departure for our research. Subsection 4.2 reviews work on problems of estimation when some state components are known exactly (the singular estimation problem) or very accurately (the almost singular estimation problem). Multiple time scales arise in the solution of almost singular estimation problems, and singular perturbation methods can be used to solve such problems. As noted in the previous section, the almost singular estimation problem is equivalent to a simple variation of one of our models. Thus, both the results and the methods are of interest. Subsection 4.3 discusses work on estimation problems with natural time scales -- that is, problems in which the time scale is introduced explicitly as a singularly-perturbed filtering problem. Although the models used in this work do not quite include the one we have described for the passive acoustic tracking problem, they are closely related and the methods of analysis are of interest.

4.2 SINGULAR AND NEARLY-SINGULAR FILTERING PROBLEMS

4.2.1 Introduction

The problem of interest in our work is characterized by a very high intensity for the measurement noise, but with a long time horizon to perform tracking, so many measurements may be taken. Ideally, we expect a tradeoff between the poor measurement quality and the number of measurements which are available. Since this is the main distinguishing feature of the problem from other nonlinear filtering problems, it is papers in this area that are most thoroughly discussed.

The problem of poor measurements over a long time horizon has received little, if any, attention in the literature. the related problem of estimation with good measurements over a very short period of time has been dealt with extensively. In addition, the dual to this estimation problem, the cheap control problem, has received significant attention.

4.2.2 Singular Estimation and Control

The limiting case of estimation in low noise (control with small input penalty) is the situation where noise intensity is zero (penalty associated with the inputs is zero). These problems are known as the singular estimation and singular control problems, respectively. They have been examined most extensively for linear Gaussian models.

For the case of linear systems with a continuous state space, Willems, Kitapci and Silverman [14] investigated the singular control problem while Schumacher [15] examines singular filtering. Both papers deal with systems that can be described in the form:

$$\dot{x}(t) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t) + Du(t)$$

where u(t) is the input in the control problem and noise process in the estimation problem.

Schumacher handles the filtering problem by breaking the state space into two subspaces, one which can be estimated exactly because part of the observations are uncorrupted by measurement noise and the other which is observed only in noise. The representation of the system becomes:

$$\dot{x}_1(t) = A_{11}x_1(t) + A_{12}x_2(t) + B_1u(t)$$

$$\dot{x}_2(t) = A_{21}x_1(t) + A_{22}x_2(t) + B_2u(t)$$

$$y(t) = C_1x_1(t) + C_2x_2(t) + Du(t) \quad .$$

$x_2(t)$ can be determined exactly from y and its derivatives and also we can construct a matrix G_1 so that

$$\dot{x}_1(t) = (A_{11}+G_1C_1)x_1(t)-G_1y(t) \quad ,$$

Note that the equation for $\dot{x}_1(t)$ driven by y(t) is of the same form as a Kalman filter and indeed this reduces to the usual Kalman filter when there are no noise-free observations.

The work of Willems et al. [14] deals with the control version of the problem. Given the state variables x(t) and control variables u(t) from

$$\dot{x}(t) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t) + Du(t) \quad ,$$

some of the controls are associated with zero penalty. Therefore, there are two subspaces for u, one in which infinite magnitude inputs such as impulses are allowed (zero cost portion) and the other where regular (finite magnitude) inputs only are possible. The state space is divided into subspaces such as the output nulling subspace, controllable subspaces, almost nulling subspaces and combinations thereof. The paper shows that the optimal control consists of a set of inputs which forces the state to jump to the subspace controllable by regular inputs. The subsequent trajectory of the state remains within the subspace.

This compares to the filtering problem of Schumacher where the uncorrupted subspace was estimated exactly and instantly, while the variables in the remaining subspace were estimated over a finite time interval, with imperfect accuracy, by linear filtering techniques.

The singular filtering problem for discrete time linear systems has been addressed by Shaked [16]. The system investigated is of the form

$$x_{i+1} = Ax_i + Bw_i$$

$$y_i = Cx_i + v_i$$

where $E[v_i v_j] = r\delta_{ij}$ and $r \neq 0$.

The paper distinguishes between the cases of uniform and nonuniform rank. A uniform rank system has the property

$$CA^iB = 0 \quad \text{for } i = 0,1,\ldots,\ell-1$$

and

$$\det|CA^\ell B| \neq 0 \quad ,$$

which simply means that there is an equal number of integrations (continuous time) or delays (discrete time) between the inputs and the outputs.

A solution is given explicitly in both cases for the steady-state values of the Kalman gain and the error covariance matrix.

4.2.3 Singularly-Perturbed and Nearly-Singular Estimation and Control

The previous subsection dealt with filtering problems that had uncorrupted measurements and control problems with inputs associated with zero penalty. A great deal of attention has been given in the literature to closely related nearly-singular problems. In these cases, the noise in the estimation problem and penalty on inputs in the control problem are small and positive but not exactly zero. The nearly singular estimation problem is discussed by Krener [17] for linear systems (and a restricted class of nonlinear systems), while the linear cheap control problem is handled by Sannuti [18]-[21].

The work by Krener deals with systems of the form:

$$dx = Axdt + B(\varepsilon)dw \qquad w, v, \text{ Brownian motion processes.}$$
$$dy = Cxdt + D(\varepsilon)dv \quad ,$$

with the matrices $B(\varepsilon)$ and $D(\varepsilon)$ being a function of the small positive parameter ε. The approach used to solve the filtering problem by Krener is to take the perturbation parameter ε out of the noise covariance matrix and incorporate it into the dynamics of the system. By assuming particular structure for $B(\varepsilon)$ and $D(\varepsilon)$, he first transforms the system above into the form

$$\begin{bmatrix} dx_F \\ dx_s \end{bmatrix} = \begin{bmatrix} A_{ff} & A_{fs} \\ A_{sf} & A_{ss} \end{bmatrix} \begin{bmatrix} x_t \\ x_s \end{bmatrix} dt + \begin{bmatrix} B_{ff}(\varepsilon) & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dw_f \\ dw_s \end{bmatrix}$$

$$\begin{bmatrix} dy_f \\ dy_s \end{bmatrix} = \begin{bmatrix} C_{ff} & 0_s \\ 0 & C_{ss} \end{bmatrix} \begin{bmatrix} x_f \\ x_s \end{bmatrix} dt + \begin{bmatrix} D_{ff}(\varepsilon) & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} d\nu_f \\ d\nu_s \end{bmatrix}$$

where x_f are the states which can be estimated quickly and y_f are the outputs which are corrupted only by a low-intensity noise process. A further assumption on $B_{ff}(\varepsilon)$ and $D_{ff}(\varepsilon)$ and a scaling of the state and output variables yields a system of the form:

$$\begin{bmatrix} d\xi_f \\ d\xi_s \end{bmatrix} = \begin{bmatrix} \varepsilon^{-1}A_{ff} & \varepsilon^{-1/2}A_{fs} \\ 0(\varepsilon^{1/2}) & A_{ss} \end{bmatrix} \begin{bmatrix} \xi_f \\ \xi_s \end{bmatrix} dt + \begin{bmatrix} \varepsilon^{-1/2} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dw_f \\ dw_s \end{bmatrix}$$

$$\begin{bmatrix} d\psi_f \\ d\psi_s \end{bmatrix} = \begin{bmatrix} \varepsilon^{-1}C_{ff} & 0 \\ 0 & C_{ss} \end{bmatrix} \begin{bmatrix} \xi_f \\ \xi_s \end{bmatrix} dt + \begin{bmatrix} \varepsilon^{1/2}I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} d\nu_f \\ d\nu_s \end{bmatrix}$$

The resulting perturbations in the dynamics of the system allow Krener to approximate the solutions to the entire Riccati equation. He solves smaller-order problems by expanding the solution in a power series and keeping only leading-order powers of ε. The solution demonstrates that there are two rates associated with the dynamics of the Riccati equation when some states are corrupted by small noise.

Krener extends the problem to the case where

$$\begin{bmatrix} dx_f \\ dx_s \end{bmatrix} = \begin{bmatrix} A_{ff} & A_{fs} \\ A_{sf} & A_{ss} \end{bmatrix} \begin{bmatrix} x_f \\ x_s \end{bmatrix} dt + \begin{bmatrix} 0 \\ \alpha_s(x_f) \end{bmatrix} dt + \begin{bmatrix} B_{ff} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dw_f \\ dw_s \end{bmatrix}$$

and shows that the optimal filter can be approximated by the linearized version, where $\alpha_s(x_f)$ is replaced by $\alpha_s(\bar{x}_f)$ and $\bar{x}_f$ is the estimate of the variables which can be quickly estimated.

The cheap control problem for linear systems has been investigated by Sannuti [18]-[21]. In [18] and [21] specifically, the linear regulator problem is considered in which the controls are penalized by an amount which is $O(\varepsilon) = O(\mu^2)$ smaller than the penalty associated with outputs of equivalent magnitude. Therefore we have a system of the form:

$$\dot{x} = Fx + Gu$$

$$y = Hx$$

$$J = \frac{1}{2} x^T S(t_f)x + \int_{t_o}^{t_f} \left(y^T A y + \mu^2 u^T B u \right) dt \quad .$$

In [22], Sannuti and Wason demonstrate how an invertible system can be put into an "almost observable" form using an invertible linear transformation. The system can be written as the following when in this form:

$$\dot{x}_0 = A_{00}x_0 + A_{01}x_1$$

$$\dot{x}_{ia} = \sum_{j=0}^{k} A_{ij}x_j + D_i u \quad , \quad i = 1,\ldots,k$$

$$\dot{x}_{ib} = \alpha_i x_i + x_{i+1} \quad , \quad i = 1,\ldots,k-1$$

$$y = Hx_1 \quad .$$

The structure generated by this form is shown in a block diagram in Fig. 4-1. The key feature to this structure is the number of integrations along each path from input to output. In general, there are paths with 1,2, ... K-1 and K integrators in series. Given the representation above and the cost function, Sannuti shows that the magnitude of the optimal controls will be $O(\mu^{-1})$. Furthermore, the high gain of $O(\mu^{-1})$ will be evenly distributed across the integrations of each path.

Therefore, for the forward path with three integrators, the inputs to the integrators will be $O(\mu^{-1})$, $O(\mu^{-2/3})$, and $O(\mu^{-1/3})$ as shown in Fig. 4-2. In the additional dynamics for x_0, the integrator input is the same order as y.

To understand the effect on the dynamics of the control law, we recognize that the integrator inputs are the derivatives of the state variables, so that the dynamics are clearly speeded up. Sannuti proceeds to scale the state variables and controls so that the new inputs to the integrators are all of the same order. For the case of a three-integration chain we can consider a simple example.

$$\begin{bmatrix} \dot{\tilde{x}}_1 \\ \dot{\tilde{x}}_2 \\ \dot{\tilde{x}}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tilde{u}$$

$$\tilde{y} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tilde{x} \qquad J = \int_0^T \left(\tilde{y}^T \tilde{y} + u^T u \mu^2 \right) dt \quad .$$

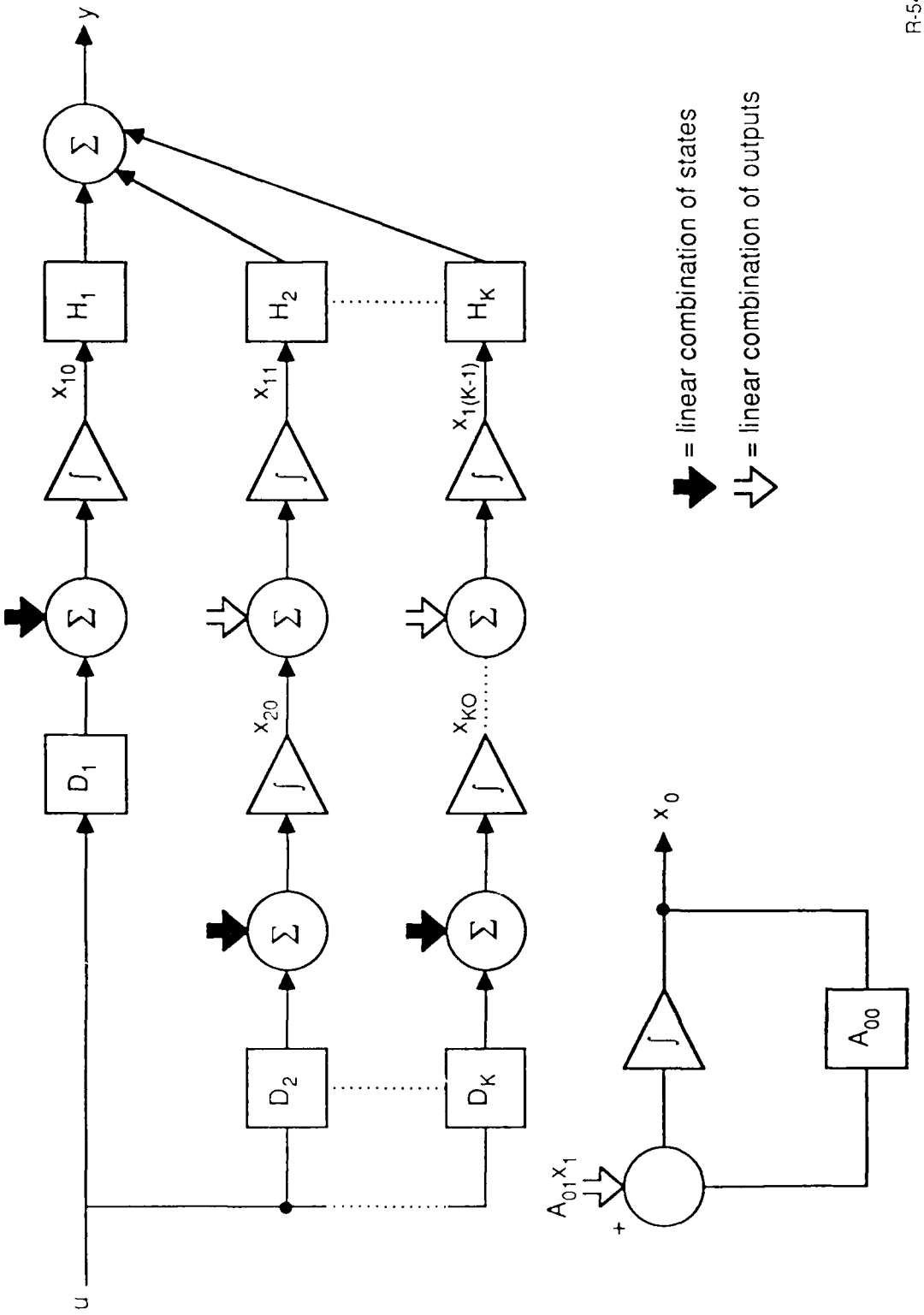


Figure 4-1. Structure of Sanmuti's "Almost Observable Form"

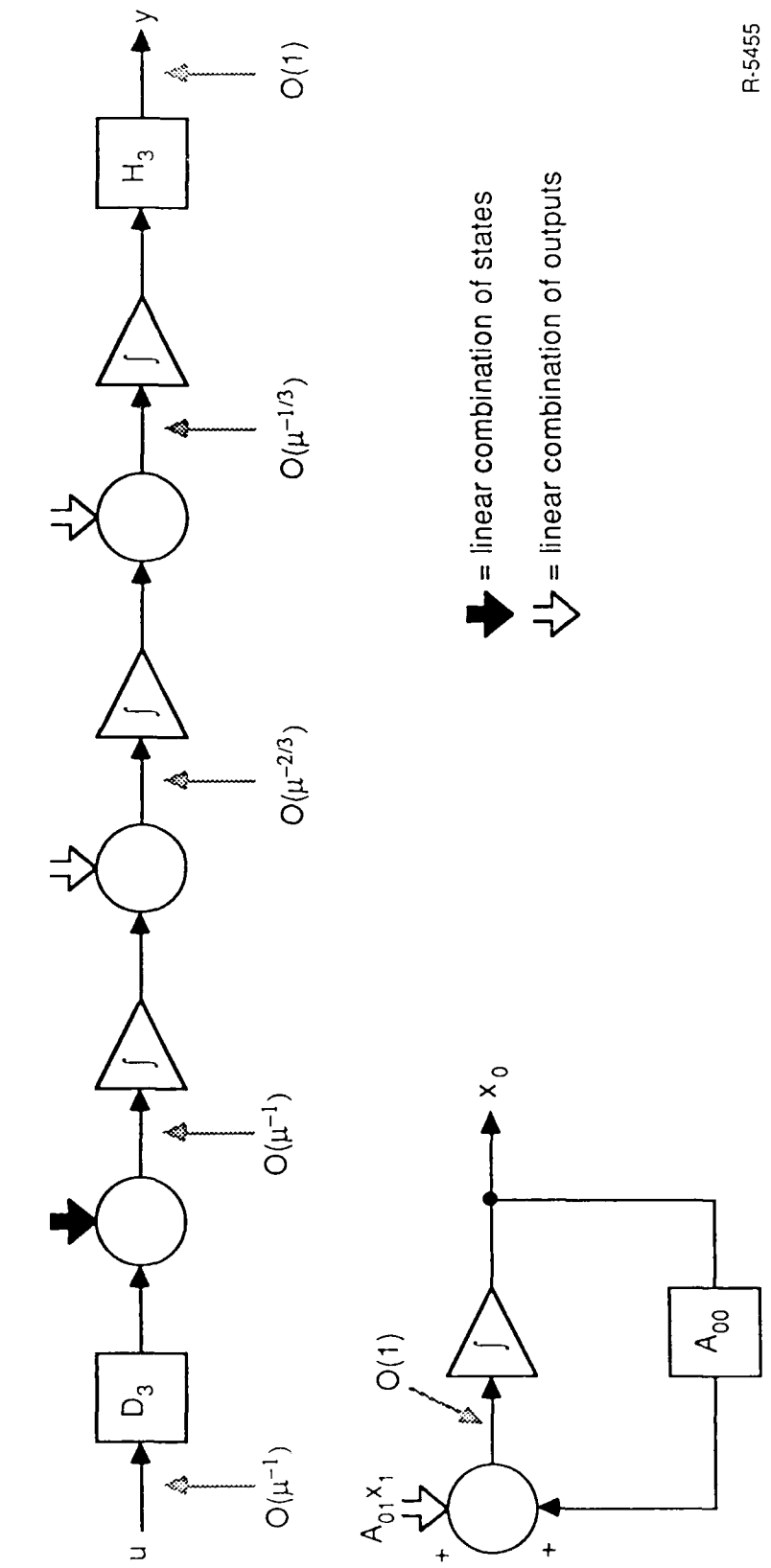


Figure 4-2. Distribution of High-Gain Optimal Input Across the Integration Chain

Under the optimal control we will have:

$$O(\tilde{u}) = O(\mu^{-1})$$

$$O(\tilde{x}_3) = O(\mu^{-2/3})$$

$$O(\tilde{x}_2) = O(\mu^{-1/3})$$

$$O(\tilde{x}_1) = O(y) = O(1) \quad .$$

To scale all the derivatives to order $O(\mu^{-1/3})$ we let

$$u = \mu\tilde{u}$$

$$x_3 = \mu^{2/3}\,\tilde{x}_3$$

$$x_2 = \mu^{1/3}\,\tilde{x}_2$$

$$x_1 = \tilde{x}_1$$

to obtain:

$$\mu^{1/3}\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} + \begin{bmatrix}0\\ 0\\ 1\end{bmatrix} u$$

$$J = \int_0^T \left(\tilde{y}^T\tilde{y} + u^T u \right) dt \quad .$$

Therefore the result of the μ^2 in the original cost function is a time scaling of $\mu^{1/3}$. In the case where there are k integrations in series, the scaling will be $\mu^{1/k}$. Therefore, for systems with more than one forward path (multiple rank) there will be multiple time scales on which system behavior occurs. For the additional system,

$$\dot{\tilde{x}} = A_{01}\tilde{x}_1 + A_{00}\tilde{x}_0$$

we let

$$x_0 = \tilde{x}_0$$

and therefore there is no time scaling effect.

Due to the duality between estimation and control problems, we can expect a similar phenomena to occur in estimation problems. The above phenomena resulted from the $\mu^2 B$ term in the penalty function. The dual condition for the estimation problem would be that

$$\text{cov}[\dot{v}(t)\dot{v}(s)] = \mu^2 R\delta(t,s)$$

which corresponds to the case of very good measurements [23, p. 270].

To determine the behavior of the estimation problem, we start by working with general control and estimation problems, and generate the dual of the cheap control problem.

CONTROL/ESTIMATION DUALITY

To obtain a general relationship between the (nearly) singular control and (nearly) singular filter problems, we need to express the systems in the general form:

CONTROL	ESTIMATION
$\dot{x} = Fx + Gu$	$\dot{x} = F^T x + H^T \omega$
$y = Hx$	$y = G^T x + \dot{v}$
$J = \frac{1}{2} x(t_f)^T S(t_f) x(t_f) + \frac{1}{2} \int_{t_0}^{t_f} y^T A y + u^T B u \; dt$	$\text{cov}(\dot{\omega}) = A = Q$ $\text{cov}(\dot{v}) = B = R$.

Therefore, given the estimation problem:

$$\dot{x} = Fx + H\dot{w} \quad ; \quad \text{cov}(\dot{w}) = I$$

$$y = Gx + \dot{v} \qquad \text{cov}(\dot{v}) = \mu^2 I \quad ,$$

we can translate into the control problem:

$$\dot{x} = F^T x + G^T u \quad .$$

$$y = H^T x$$

$$J = \frac{1}{2} \int_{t_0}^{t_f} y^T y + \mu^2 u^T u \quad .$$

The transformation described by Sannuti could then be performed so that:

$$\dot{x}_0 = A_{00} x_0 + A_{01} x_1$$

$$\dot{x}_{ia} = \sum_{j=0}^{k} A_{ij} x_j + D_i u \quad , \quad i = 1,\ldots,k$$

$$\dot{x}_{ib} = \alpha_i x_1 + x_{i+1} \quad , \quad i = 1,\ldots,k-1$$

$$y = H x_1$$

$$J = \frac{1}{2} \int_{t_0}^{t_f} y^T y + \mu^2 u^T u \quad .$$

Translating back into the estimation context we obtain:

$$\dot{x}_0 = A_{00}^T x_0 + \sum_{i=1}^{k} A_{io}^T x_{ia}$$

$$\dot{x}_j = \sum_{i=1}^{k} A_{ij}^T x_{ia} + x_{(j-1)b}$$

$$\dot{x}_1 = A_{01}^T x_0 + \sum_{j=1}^{k} A_{j1}^T x_{ia} + \sum_{i=1}^{k=1} \alpha_i x_{ib} + H^T u$$

$$y = \sum_{i=1}^{k} D_i^T x_{ia} + \dot{v}$$

$$\text{cov}\left((\dot{v}(t),\dot{v}(s)\right) = B\delta(t,s) \quad .$$

The original equations for the control problem implied the structure of Fig. 4-1, while those for the filtering problem imply the structure of Fig. 4-3. Therefore, we find that in the dual system, the form of the transformed structure is the same (chains of integrators), with process noise replacing the inputs and measurement noise added to the outputs.

The only minor differences are that the input and output matrices have reversed position and the state variables have assumed a transposed position in the structure. However, since the variable labeling is arbitrary, we can expect results for the filtering problem which are identical in form to those obtained for the control problem.

To demonstrate the implications of Sannuti's results for an estimation problem, consider a simple example. Given that the special form required by Sannuti is quite similar for the estimation problem, we can formulate a problem that is already in the required form. Assume for simplicity that we have a similar system to the one considered in the control example:

$$\dot{\tilde{x}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tilde{x} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \dot{w}_1 = A\tilde{x} + B\dot{w}_1$$

$$\tilde{y} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tilde{x} + \mu^2\dot{v} = C\tilde{x} + \mu^2\dot{v}$$

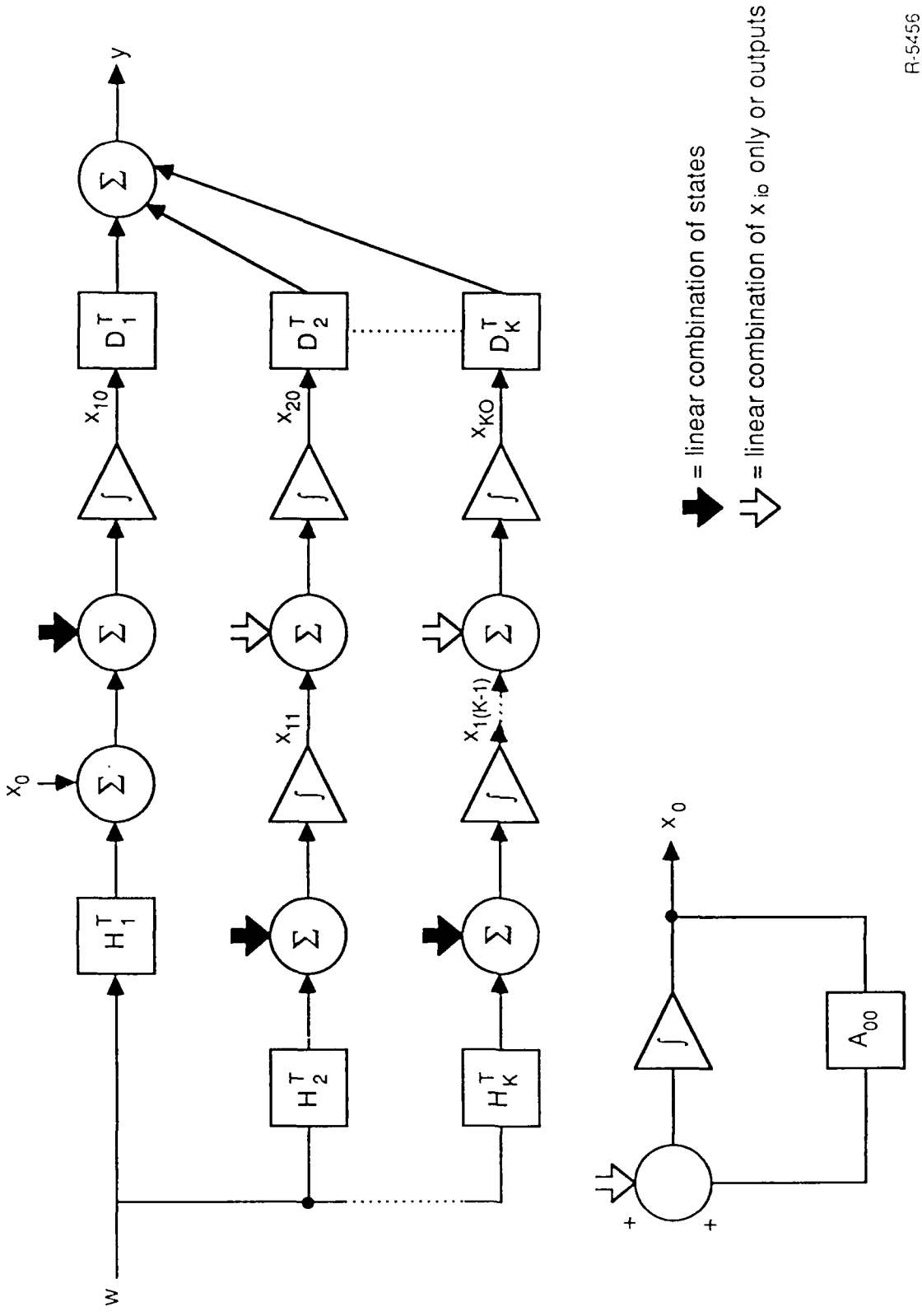


Figure 4-3. Structure Implies by Estimation Dual of Almost Observable Form

We can write the filtering equations:

$$\dot{\tilde{\hat{x}}} = A\tilde{\hat{x}} + \tilde{P}C^T\mu^{-2}(\tilde{y}-C\tilde{\hat{x}})$$

$$\dot{\tilde{P}} = A\tilde{P} + \tilde{P}A^T + BB^T - \tilde{P}C^TC\tilde{P}\mu^{-2}$$

which for our example become:

$$\dot{\tilde{\hat{x}}}_1 = \tilde{\hat{x}}_2 + \mu^{-2}\tilde{P}_{11}(\tilde{y} - \tilde{\hat{x}}_1)$$

$$\dot{\tilde{\hat{x}}}_2 = \tilde{x}_3 + \mu^{-2}\tilde{P}_{12}(\tilde{y} - \tilde{\hat{x}}_1)$$

$$\dot{\tilde{\hat{x}}}_3 = \mu^{-2}\tilde{P}_{13}(\tilde{y} - \tilde{\hat{x}}_1) \quad .$$

In the steady state,

$$\tilde{P}_{11} = 2\mu^{5/3} \qquad \tilde{P}_{22} = 3\mu$$

$$\tilde{P}_{12} = 2\mu^{4/3} \qquad \tilde{P}_{23} = 2\mu^{2/3}$$

$$\tilde{P}_{13} = \mu \qquad \tilde{P}_{33} = 2\mu^{1/3}$$

which result from the Riccati equations:

$$\dot{\tilde{P}}_{11} = 2\tilde{P}_{12} - \mu^{-2}\tilde{P}_{11}^2$$

$$\dot{\tilde{P}}_{12} = \tilde{P}_{22} + \tilde{P}_{13} - \mu^{-2}\tilde{P}_{11}\tilde{P}_{12}$$

$$\dot{\tilde{P}}_{13} = \tilde{P}_{23} - \mu^{-2}\tilde{P}_{11}\tilde{P}_{13}$$

$$\dot{\tilde{P}}_{22} = 2\tilde{P}_{23} - \mu^{-2}\tilde{P}_{12}^2$$

$$\dot{\tilde{P}}_{23} = \tilde{P}_{33} - \mu^{-2}\tilde{P}_{12}\tilde{P}_{13}$$

$$\dot{\tilde{P}}_{33} = 1 - \mu^{-2}\tilde{P}_{13}^2 \quad .$$

Now consider the transformation:

$$x_1 = \mu^{-5/6}\tilde{x}_1$$

$$x_2 = \mu^{-1/2}\tilde{x}_2 \qquad y = \mu^{-5/6}\tilde{y} \quad .$$

$$x_3 = \mu^{-1/6}\tilde{x}_3$$

Then the Riccati equations become:

$$\mu^{1/3}\begin{bmatrix}\dot{P}_{11}\\ \dot{P}_{12}\\ \dot{P}_{13}\\ \dot{P}_{22}\\ \dot{P}_{23}\\ \dot{P}_{33}\end{bmatrix} = \begin{bmatrix}2P_{12} - P_{11}^2\\ P_{22} + P_{13} - P_{11}P_{12}\\ P_{23} - P_{11}P_{13}\\ 2P_{23} - P_{12}^2\\ P_{33} - P_{12}P_{13}\\ 1 - P_{13}\end{bmatrix}$$

$$\tilde{P}_{11} = \mu^{5/3}P_{11}$$

$$\tilde{P}_{12} = \mu^{4/3}P_{12}$$

$$\tilde{P}_{13} = \mu P_{13}$$

$$\tilde{P}_{22} = \mu P_{22}$$

$$\tilde{P}_{23} = \mu^{2/3}P_{23}$$

$$\tilde{P}_{33} = \mu^{1/3}P_{33}$$

Solving the algebraic Riccati equations, we get

$$P_{11} = 2 \qquad P_{23} = 3$$

$$P_{12} = 2 \qquad P_{23} = 2$$

$$P_{13} = 1 \qquad P_{33} = 2$$

which upon transformation back to $\tilde{P}$ yields our original result.

The filtering equations become:

$$\mu 1/3 \begin{bmatrix} \dot{\hat{x}}_1 \\ \dot{\hat{x}}_2 \\ \dot{\hat{x}}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{bmatrix} + \begin{bmatrix} P_{11} \\ P_{12} \\ P_{13} \end{bmatrix} (y-\hat{x}_1) \quad .$$

This result clearly resembles that obtained for the control problem (time-scaling of the order $\mu^{1/3}$).

From this new form of equations we can easily see the effects of the low measurement noise.

1. The dynamic behavior of the Riccati equation is "speeded up," thereby indicating that the filter reaches steady-state much more quickly.

2. The filter dynamics themselves are speeded up.

3. The increased rate of filter convergence is given by a factor $\mu^{1/3}$ with three integrations. This would be $\mu^{1/k}$ for k integrations, paralleling the cheap control case.

4. The ending steady-state covariances are decreased much more for state variables that are close to the output rather than the input (see Fig. 4-4). Therefore, the difference in covariance of the process and measurement noise intensities is spread evenly across the integrators, so low measurement noise may not be extremely useful if we are estimating state variables in a long sequence of integrations.

5. For the opposite case, when μ is very large instead of very small, the trend is reversed. We will be capable, comparatively, of estimating variables close to the process, while variables that are close to the measurement in the chain may not be possible to estimate.

6. For the reversed case (μ large), the convergence of the covariance (and filter) to the steady state will be very slow.

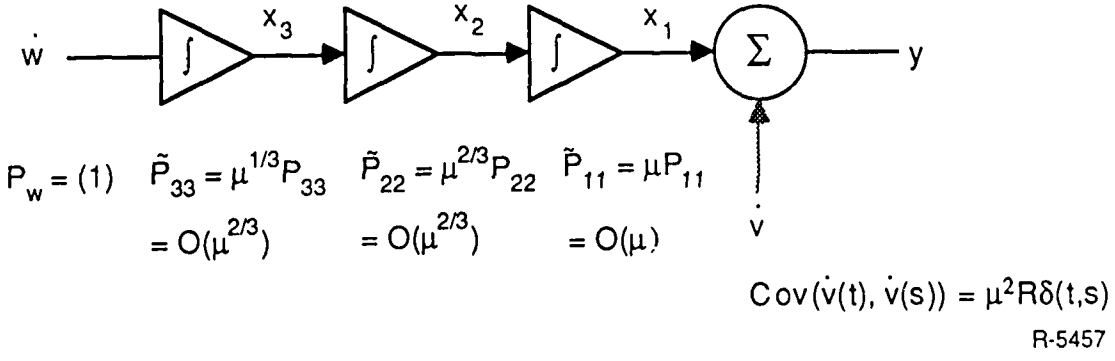


Figure 4-4. Effect of Low Measurement Noise on Steady-State Error Covariances

NEARLY-SINGULAR NONLINEAR ESTIMATION

The filtering problem for this type of model is discussed by Katzur, Bobrovsky, and Schuss [24]. A system of the form:

$$\begin{aligned} dx_t &= m(x_t)dt + \sigma dw_t \\ dy_t &= h(x_t)dt + \varepsilon dv_t \end{aligned} \qquad \text{(linear observations)}$$

is discussed, where $h(x_t)$ is linear in x_t, w_t and v_t are unit intensity Brownian motion processes and ε is a small positive parameter. The propagation of the conditional density function in the form of Kushner's equation is discussed. This yields an equation of the form:

$$dp = A^*pdt + \frac{p}{\varepsilon^2}(h-\hat{h})\left(dy - \hat{h}(x_t)\right)dt$$

which is approximated using a power series expansion in ε. A similar procedure is used for Zakai's equation for the unnormalized conditional distribution in the case where $h(x_t) = x_t$ (dealing with scalars).

An estimate for x (leading term) is then obtained as the solution to the differential equation

$$d\hat{x} = -\sigma\frac{\hat{x}}{\varepsilon} + \frac{\sigma}{\varepsilon}dy = \frac{\sigma}{\varepsilon}(dy-\hat{x}) \quad .$$

Additional terms are obtained in higher orders of ε to improve accuracy. These additional terms are found as functions of x_0^* where x_0^* is the solution to the differential equation above. This yields, to first order in ε,

$$\hat{x} = x_0^* + \frac{\varepsilon}{\sigma}m(x_0^*) + \ldots$$

The advantages of this formulation is that a fast analog processor can perform the averaging process required in the calculation of x_0^*. After this smoothing integration, a slower sampling can be performed, followed by the nonlinear approximation above. Therefore, this technique naturally yields an estimator structure with a fast front end" and a slow "back end" processor.

4.3 FILTERING WITH DYNAMICS POSSESSING NATURAL TIME SCALES

A second type or perturbed filtering problem which is of interest is the case where the perturbation parameter is in the dynamics of the system instead of the noise magnitudes. This problem and its control dual have been investigated by Chow and Kokotovic [25], Haddad [26], and Marchetti [27],[28]. In these cases the system displays behavior on multiple time scales before the noise levels or control penalties are considered.

4.3.1 Continuous-State Problems

Chow and Kokotovic [25] analyze a control problem which is nonlinear and displays behavior on two time scales. The model employed is of the form

$$\dot{x} = a_1(x) + A_1(x)z + B_1(x)\,u$$

$$\varepsilon\dot{z} = a_2(x) + A_2(x)z + B_2(x)\,u \quad .$$

$$J = \int_0^\infty \left[P(x)+S'(x)z + z'Q(x)z + u'R(x)u\right] dt \quad .$$

In their analysis, they attempt to obtain an approximate, but lower dimension solution to this problem. The approach that is used is to initially set $\varepsilon=0$, solve for z and then determine the optimal slow control for the reduced-order system. The fast subproblem is then solved using the fast

subproblem is then solved using the fast dynamics and the solution to the slow control problem. The two controls are then added together to obtain:

$$u_c = u_s + u_f \quad .$$

where u_s is the optimal control for $\varepsilon=0$ and u_f is a correction term required because $\varepsilon\neq 0$. The dual estimation problem is quite similar. The extra difficulty that arises for the estimation problem is that we also have white noise present, which is fast when viewed at any time scale. The dual problem is, however, discussed by Haddad [26] with results similar to those for the control problem.

Haddad considers a system of the form:

$$\begin{aligned} \dot{x} &= A_1 x + A_{12} z + B_1 u \\ \varepsilon \dot{z} &= A_{21} x + A_{22} z + B_2 u \\ y &= C_1 x + C_2 z + v \quad . \end{aligned}$$

His first step is to perform a linear transformation to separate the fast and slow parts of the state to obtain:

$$\begin{aligned} \dot{\eta} &= A_0 \eta + B_0 u \\ \varepsilon \dot{\xi} &= A_2 \xi + B_2 u \qquad & \eta(t_0) &= \eta_0 \\ \dot{y} &= C_0 \eta + C_2 \xi + v \qquad & \xi(t_0) &= \xi_0 \quad . \end{aligned}$$

As in Chow and Kokotovic's solution to the control problem, Haddad splits the solution into two parts. The first part is obtained by setting $\varepsilon=0$ and solving for ξ in terms of u, thereby assuming for this portion that ξ is a white-noise process. The resulting "slow" solution is shown to be valid to $O(\varepsilon)$ for $t > t_0 + \mu$, where μ is a small positive number. For the interval

$t_0 < t < t_0 + \mu_1$, an additional portion, known as the boundary layer, must be added to the solution.

The term "boundary layer" refers to an additional part of the solution which is an adjustment for the facts that we have initial conditions and the ξ process is not actually white noise.

The resultant filter is given by:

$$\dot{\hat{n}} = A_0\hat{n} + (P_1C_0' + P_{12}C_2')R^{-1}(y - C_0\hat{n} - C_2\hat{\xi})$$

$$\mu\dot{\hat{\xi}} = A_2\hat{\xi} + (\varepsilon P_{12}C_0' + P_2C_2')R^{-1}(y - C_0\hat{n} - C_2\hat{\xi}) \quad .$$

P_2 and P_{12} are given by

$$P_2(t) = \overline{P_2(t)} + \widetilde{P_2(\theta)} + O(\varepsilon)$$

$$P_{12}(t) = \overline{P_{12}(t)} + \widetilde{P_{12}(\theta)} + O(\varepsilon)$$

where $\theta = t/\varepsilon$ and $\widetilde{P_2(\theta)}$, $\widetilde{P_{12}(\theta)}$ are corrections for boundary layer effects. The slow mode filter may be implemented without the correction terms and still be valid for all $t > t_0$ because the filter dynamics are slow and therefore the boundary layer time interval will have negligible effect on its output.

4.3.2 Discrete-State Problems

Marchetti [27] deals with discrete-state space problems, specifically Markov chains. The system dynamics in these problems are singularly perturbed. The systems dealt with are of the form:

$$dp_t^{\varepsilon} = (B+\varepsilon A)p_t^{\varepsilon}$$

$$dy_t = h(x_t^{\varepsilon})dt + dw_t^{\varepsilon}$$

where p_t^{ϵ} is an n-dimensional column vector of probabilities, $(B+\epsilon A)$ is a state transition matrix, w^{ϵ} is a standard Brownian motion process and $h(x_t^{\epsilon})$ is a function of the state. The fact that the transition matrix is $B+\epsilon A$, will yield multiple time-scale behavior for the system, given that B satisfies certain conditions on its eigenvalues. Marchetti derives formulas for propagating the Zakai equation by

1. Writing p_t^{ϵ} as a power series in ϵ,

$$p_t^{\epsilon} = \sum_{k=0}^{\infty} p_k(t)\epsilon^k \quad .$$

2. Substituting in the Zakai equation and equating power of ϵ.

3. Writing the equations above in aggregate and decentralized forms, thereby reducing the order of the systems of differential equations.

Using the second step we can approximate the probability mass function for the states of the system to any order in ϵ that we desire; however, the approximation will only be valid on an interval $[0,T]$. In [28] a variation on this is developed which produces approximations over intervals of length $[0,T/\epsilon]$.

SECTION 5

FILTER ARCHITECTURES FOR TIME-SCALE APPROXIMATIONS

5.1 COMPARISON OF FILTERING TECHNIQUES TO THE REQUIREMENTS OF OUR PROBLEM

From our original discussion we are interested in the estimation problem for the system

$$dx = A_1(x) + B_1(\varepsilon)\, dw_1$$

$$\varepsilon dz = A_2(x) + B_2(\varepsilon)\, dw_2$$

$$(\varepsilon)dy = C(x_1, z) + D(\varepsilon)\, dv \quad .$$

The noise terms have magnitudes which are functions of ε and therefore the system is in the nearly singular category (in addition to a perturbation in the dynamics). In particular, we are interested in the case of poor measurements so we might set $D(\varepsilon) = \sqrt{\varepsilon}\ I$, and $O(B_1(\varepsilon)) = O(1)$, $O(B_2(\varepsilon)) = O(\sqrt{\varepsilon})$.

The problem considered by Katzur et al. [24] falls into the same problem category, but deals only with the scalar case and a system with one time scale. In addition, it deals with small noise instead of large noise. Finally, the nonlinearities occur only in the state dynamics, not the observations.

The work by Haddad [26] is similar to our problem except that ours is nonlinear and he does not assume ε dependence of the noise magnitudes. In addition, his measurements are a function of both the fast and slow processes, whereas we are measuring only the fast process.

Krener [17] does not deal with more than one natural time scale and assumes that the ε dependence of the noise produces small noise instead of large noise. Chow and Kokotovic [25] deal with a system with natural time scales and nonlinearities. However, the problem does not contain ε dependencies in the cost function which would become noise magnitudes in the filtering dual.

Sannuti does not assume natural time scales for the system, but does deal with time scales once they arise from perturbation parameters in the penalty function. The dual of his assumptions, however, are that the noise intensity is low.

Finally, Marchetti assumes perturbation parameters in the system dynamics. The noise is also assumed to be bad in the case of [28]. In [27] he handles an approximation that is valid only on [0,T], but [28] improves this to the interval $[0,1/\varepsilon]$.

In general, the exact problem that we are interested in has not been handled in the literature due to a combination of the following features:

1. Measurements are nonlinear functions of the state.
2. Perturbations exist in both the dynamics and the noise magnitudes.
3. We are interested in the high-noise case, not the low-noise case.

Regardless of the exact applicability of the techniques covered in the literature, we can examine the structures of the processors that are implied by each of the techniques (or corresponding duals).

5.2 FILTER ARCHITECTURES OF EXISTING TECHNIQUES

5.2.1 Fast Front End, Slow Back End With Feed-Forward and Feedback (Fig. 5-1)

This structure was motivated by the structures of existing processors. Not all existing structures have the feedback present, and depending on the problem conditions, it may or may not be very useful.

5.2.2 Slow Processor with Correction (Fig. 5-2)

This structure is motivated by the work of Chow and Kokotovic [25] and Haddad [26]. In both cases, the slow processor is the processor that would be obtained if $\varepsilon=0$, with the fast dynamics assumed arbitrarily fast. The slow results are then used, in combination with the model of the fast dynamics, to determine the estimates of the fast variables.

5.2.3 Fast Estimates Followed by Slow (Fig. 5-3)

This structure is motivated by problems with ε in the cost function or noise magnitude (Schumacher, Krener, Willems). The parameter is taken into the system dynamics, resulting in a set of variables that may be estimated quickly. Once these estimates are determined (i.e., steady state is reached) they can be used to help in the estimation of variables that are corrupted by greater noise.

5.2.4 Steady-State Solution with Boundary Layer (Fig. 5-4)

This structure was motivated by Sannuti and the cheap control problem. In this case a control was found which was valid on $[0,\infty]$, but a correction had to be made over the initial time interval $[0,\varepsilon]$ to correct for the initial conditions. This correction was described as a "boundary layer."

5.2.5 Multiple Lower-Order Filter (Fig. 5-5)

This structure was motivated by Marchetti's paper on the propagation of the Zakai equation for a Markov chain. Multiple lower-order filters perform calculations which are combined to determine an approximate result for the normalized probability mass function.

5.2.6 Fast Analog Processor Followed by Slow Digital Processor (Fig. 5-6)

This structure was proposed by Katzur et al. In [24] they showed that for a nonlinear system and small measurement noise, a linear analog filter could reduce the sample frequency prior to nonlinear manipulations.

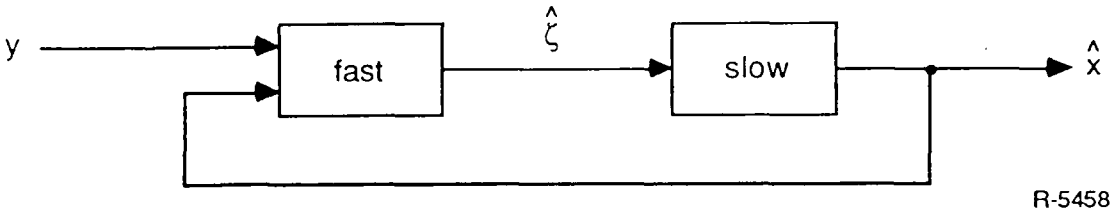


Figure 5-1. Fast Front End, Slow Back End with Feed Forward and Feedback

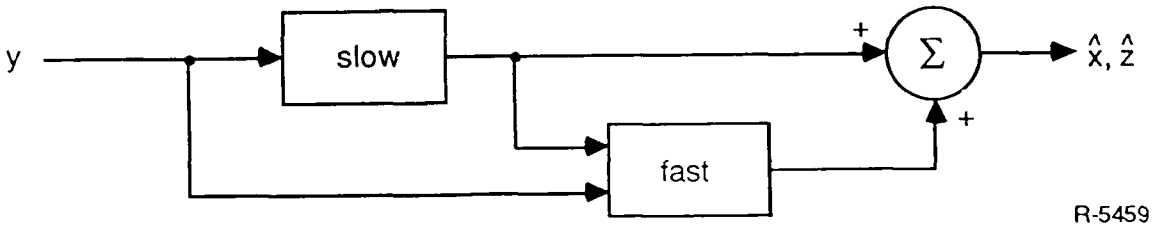


Figure 5-2. Slow Processor with Correction

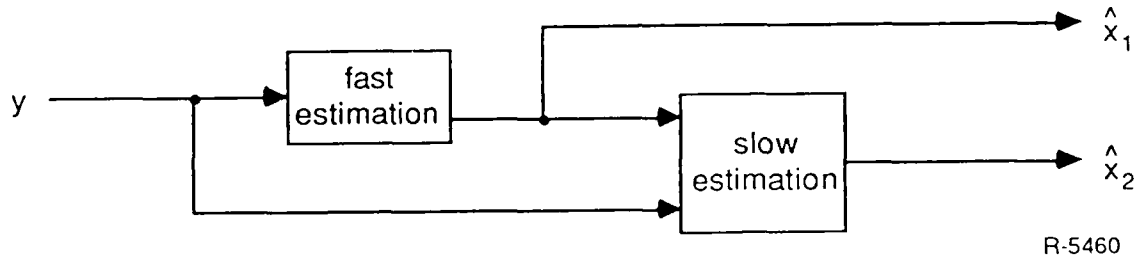


Figure 5-3. Fast Estimates Followed by Slow

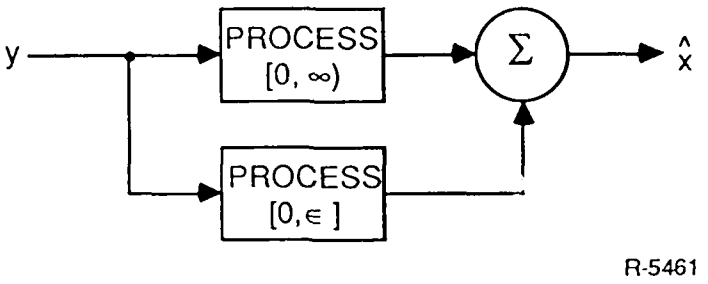


Figure 5-4. Steady-State Solution with Boundary Layer

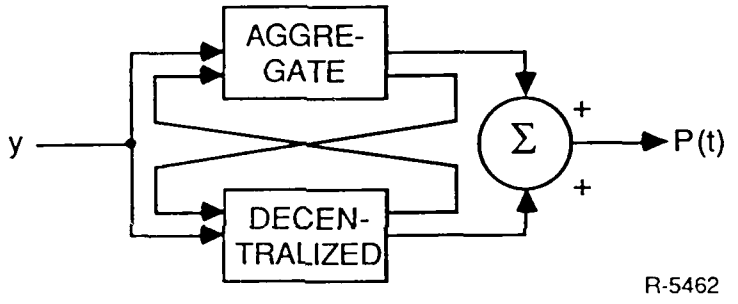


Figure 5-5. Multiple Lower-Order Filters

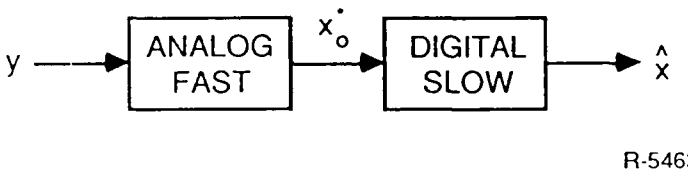


Figure 5-6. Fast Analog Processor Followed by Slow Digital Processor (Nonlinear Operations)

SECTION 6

FILTERING PROBLEM FORMULATIONS

In this section we describe both the specific nonlinear filtering problems selected for analysis and the conjectured behavior and filter architecture that these models suggest. The key features that we wished to capture in the problems chosen for study are:

- The system models should possess both fast and slow dynamics
- Direct measurements of only the fast variables are made, and these measurements may be of poor quality
- The principal objective is to estimate the slow variables
- The problems chosen should be as simple as possible in order to facilitate analysis and our ability to gain insight into the character of such estimation problems.

As discussed previously, the choice of these features was motivated by desire to capture some of the critical aspects of passive acoustic tracking problems.

6.1 EXAMPLES SELECTED FOR ANALYSIS

In this section we describe two closely related examples that have formed the focus of our detailed study. Both involve estimation for a particular finite-state Markov process possessing two time scales.

6.1.1 Measurements Corrupted by White Noise

In this problem we are interested in estimating the state of the 4-state, continuous-time finite state Markov process $\rho(t)$ whose transition behavior is depicted in Fig. 6-1. Here ε is a small parameter. This process spends most

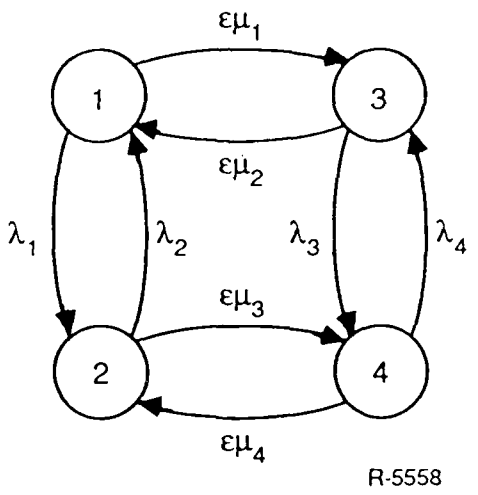


Figure 6-1. A 4-State Markov Process

of its time jumping between 1 and 2 or between 3 and 4 (the fast dynamics) and occasionally jumps between the left {1,2} and the right {3,4} (the slow dynamics). We are ultimately interested in tracking the slow behavior given noisy measurements of only the fast dynamics. The measurement model we use to capture this is

$$dy(t) = g(\varepsilon)\, h\big(\rho(t)\big)dt + dv(t) \tag{6-1}$$

where $v(t)$ is a standard Brownian motion, independent of $\rho(t)$. The quantity $g(\varepsilon)$ is a function of ε which is used to model the relative quality of our measurements. Its magnitude relative to ε, which directly controls the SNR of the measurements, is of central importance in our analysis. Finally, to guarantee that our measurements provide us only with direct information about the fast dynamics, we assume that

$$\begin{aligned} h(1) &= h(3) = \alpha \\ h(2) &= h(4) = \beta \end{aligned} \tag{6-2}$$

i.e., that our measurements indicate only if the process is on the top {1,3} or the bottom of {2,4}.

6.1.2 Perfect Measurements with Small Differences in Rates

In this second model we consider the same 4-state process in Fig. 6-1, but in this case we assume that we have perfect knowledge of whether the process is on the top {1,3} or the bottom {3,4}. In order to capture the feature that the measurements contain only weak, indirect information about left-right behavior, we assume in this case that

$$\begin{aligned} \lambda_3 &= \lambda_1 + \alpha\, g(\varepsilon) \\ \lambda_4 &= \lambda_2 + \beta\, g(\varepsilon) \quad . \end{aligned} \tag{6-3}$$

6.2 CONJECTURED ASYMPTOTIC BEHAVIOR AND FILTER ARCHITECTURES

In this subsection we perform some initial analysis of the nonlinear filtering problems introduced in the preceding subsection. The end results of these analyses are several approximate filter structure whose properties and performance versus the optimal solution are explored in subsequent sections.

6.2.1 The White Measurement Noise Model

Let

$$p_i(t) = \Pr[\rho(t) = i \mid y(\tau),\ \tau<t] \tag{6-4}$$

Then from standard results in nonlinear filtering [13], we have that

$$dp_1(t) = [(-\lambda_1 - \varepsilon\mu_1)p_1(t) + \lambda_2\, p_2(t) + \varepsilon\mu_2\, p_3(t)]dt + g(\varepsilon)[\alpha - \hat{h}(t)]\, p_1(t)[dy(t) - g(\varepsilon)\, \hat{h}(t)dt] \tag{6-5a}$$

$$dp_2(t) = [\lambda_1\, p_1(t) + (-\lambda_2 - \varepsilon\mu_3)\, p_2(t) + \varepsilon\mu_4\, p_4(t)]dt + g(\varepsilon)[\beta - \hat{h}(t)]\, p_2(t)[dy(t) - g(\varepsilon)\, \hat{h}(t)dt] \tag{6-5b}$$

$$dp_3(t) = [\varepsilon\mu_1\, p_1(t) + (-\lambda_3 - \varepsilon\mu_2)p_3(t) + \lambda_4\, p_4(t)]dt + g(\varepsilon)[\alpha - \hat{h}(t)]\, p_3(t)[dy(t) - g(\varepsilon)\, \hat{h}(t)dt] \tag{6-5c}$$

$$dp_4(t) = [\varepsilon\mu_3\, p_2(t) + \lambda_3\, p_3(t) + (-\lambda_4 - \varepsilon\mu_4)p_4(t)]dt + g(\varepsilon)[\beta - \hat{h}(t)]\, p_4(t)[dy(t) - g(\varepsilon)\, \hat{h}(t)dt] \tag{6-5d}$$

where

$$\hat{h}(t) = \alpha[p_1(t) + p_3(t)] + \beta[p_2(t) + p_4(t)] \quad . \tag{6-6}$$

As in Marchetti's work, it is quite useful to transform coordinates to highlight explicitly the aggregate variables of interest. In this example, one can give simple explanations for the several variables that arise. Specifically, let

$$p^L(t) = p_1(t) + p_2(t) \tag{6-7a}$$

$$p^R(t) = p_3(t) + p_4(t) \tag{6-7b}$$

$$\delta_1(t) = p_1(t) - \frac{\lambda_2}{\lambda_1 + \lambda_2}\, p^L(t) \tag{6-7c}$$

$$\delta_2(t) = p_3(t) - \frac{\lambda_4}{\lambda_3 + \lambda_4}\, p^R(t) \quad . \tag{6-7d}$$

Here $p^L(t)$ and $p^R(t)$ are the probabilities of being on the left and right, respectively. Also, $\delta_1(t)$ is the deviation between the exact probability of being in state 1 and its approximation obtained by multiplying the probability of being on the left by the steady-state probability of being on the top given the process is on the left and no left-right transitions can occur. An analogous interpretation can be given to $\delta_2(t)$. Note that one might expect $\delta_1(t)$ and $\delta_2(t)$ to be small since the process essentially can reach steady-state between left-right transitions.

If we transform variables according to Eq. 6-7 and eliminate $p^R(t)$ by replacing it by $1 - p^L(t)$, we obtain the following set of three coupled stochastic differential equations for the exact optimal nonlinear filter:

$$dp^L(t) = \varepsilon[\gamma_2 - (\gamma_1 + \gamma_2)p^L(t)]dt + \varepsilon[-\eta_1\ \delta_1(t) + \eta_2\ \delta_2(t)]dt + g(\varepsilon)[f_L - \hat{h}(t)]\ p^L(t)d\nu(t) + g(\varepsilon)\ \Delta\delta_1(t)d\nu(t) \tag{6-8a}$$

$$d\delta_1(t) = -\Lambda_1\ \delta_1(t)dt + \varepsilon[\eta_5 - \eta_3\ \delta_1(t) + (\eta_4 - \eta_5)p^L(t) + \eta_6\ \delta_2(t)]dt + g(\varepsilon)[\xi_1\ p^L(t) + \xi_2\ \delta_1(t) - \hat{h}(t)\ \delta_1(t)]d\nu(t) \tag{6-8b}$$

$$d\delta_2(t) = -\Lambda_2\ \delta_2(t)dt + \varepsilon[\eta_8 - \eta_7\ \delta_2(t) + (\eta_{10} - \eta_8)p^L(t) + \eta_9\ \delta_1(t)]dt + g(\varepsilon)[\xi_3 - \xi_3\ p^L(t) + \delta_4\ \delta_2(t) - \hat{h}(t)\ \delta_2(t)]d\nu(t) \tag{6-8c}$$

where

$$d\nu(t) = dy(t) - g(\varepsilon)\ \hat{h}(t)dt \tag{6-9a}$$

$$\hat{h}(t) = f_R + (f_L - f_R)\ p^L(t) + \Delta[\delta_1(t) + \delta_2(t)] \tag{6-9b}$$

and the various constants appearing in Eqs. 6-8 and 6-9 are

$$\Delta = \alpha - \beta \qquad \Lambda_1 = \lambda_1 + \lambda_2$$

$$\Lambda_2 = \lambda_3 + \lambda_4 \qquad \eta_1 = \mu_1 - \mu_3$$

$$\eta_2 = \mu_2 - \mu_4 \qquad \eta_3 = \frac{\mu_1 \lambda_1 + \mu_3 \lambda_2}{\lambda_1 + \lambda_2}$$

$$\eta_4 = \frac{\lambda_1 \lambda_2(\mu_3 - \mu_1)}{(\lambda_1 + \lambda_2)^2} \qquad \eta_5 = \frac{\mu_2 \lambda_1 \lambda_4 - \mu_4 \lambda_2 \lambda_3}{(\lambda_1 + \lambda_2)(\lambda_3 + \lambda_4)}$$

$$\eta_6 = \frac{\mu_2 \lambda_1 + \mu_4 \lambda_2}{\lambda_1 + \lambda_2} \qquad \eta_7 = \frac{\mu_2 \lambda_3 + \mu_4 \lambda_4}{\lambda_3 + \lambda_4}$$

$$\eta_8 = \frac{\lambda_3 \lambda_2(\mu_4 - \mu_2)}{(\lambda_3 + \lambda_4)^2} \qquad \eta_9 = \frac{\mu_1 \lambda_3 + \mu_3 \lambda_4}{\lambda_3 + \lambda_4} \qquad (6\text{-}10)$$

$$\eta_{10} = \frac{\mu_1 \lambda_2 \lambda_3 - \mu_3 \lambda_1 \lambda_4}{(\lambda_1 + \lambda_2)(\lambda_3 + \lambda_4)} \qquad \xi_1 = \frac{\lambda_1 \lambda_2(\alpha - \beta)}{(\lambda_1 + \lambda_2)^2}$$

$$\xi_2 = \frac{\alpha \lambda_1 + \beta \lambda_2}{\lambda_1 + \lambda_2} \qquad \xi_3 = \frac{\lambda_3 \lambda_4(\alpha - \beta)}{(\lambda_3 + \lambda_4)^2}$$

$$\xi_4 = \frac{\alpha \lambda_3 + \beta \lambda_4}{\lambda_3 + \lambda_4} \qquad \gamma_1 = \frac{\lambda_2 \mu_1 + \lambda_1 \mu_3}{\lambda_1 + \lambda_2}$$

$$\gamma_2 = \frac{\lambda_4 \mu_2 + \lambda_3 \mu_s}{\lambda_3 + \lambda_4} \qquad f_L = \frac{\lambda_1 \beta + \lambda_2 \alpha}{\lambda_1 + \lambda_2}$$

$$f_R = \frac{\lambda_3 \beta + \lambda_4 \alpha}{\lambda_3 + \lambda_4}$$

Note that the innovation $d\nu(t)$ in Eq. 6-5 is a standard Brownian motion.

Since we will be focusing on the estimation of the slow left-to-right transitions, it is useful to change time scales to that at which these transitions occur, i.e., we will make a change of scale of the form

$$t_{old} = \frac{t_{new}}{\varepsilon} \tag{6-11}$$

Performing this change of scale on Eq. 6-8, with care taken in accounting for the quadratic variation of the innovations, we obtain the following

$$dp^L(t) = [\gamma_2 - (\gamma_1 + \gamma_2)p^L(t)]dt + [\eta_2\ \delta_2(t) - \eta_1\ \delta_1(t)]dt + \frac{g(\varepsilon)}{\sqrt{\varepsilon}}[f_L - \hat{h}(t)]p^L(t)\ dw(t) + \frac{g(\varepsilon)}{\sqrt{\varepsilon}}\Delta\delta_1(t)\ dw(t) \tag{6-12a}$$

$$d\delta_1(t) = -\frac{\Lambda_1}{\varepsilon}\delta_1(t)dt + [\eta_5 - \eta_3\ \delta_1(t) + (\eta_4 - \eta_5)p^L(t) + \eta_6\ \delta_2(t)]dt + \frac{g(\varepsilon)}{\sqrt{\varepsilon}}[\xi_1\ p^L(t) + \xi_2\ \delta_1(t) - \hat{h}(t)\ \delta_1(t)]dw(t) \tag{6-12b}$$

$$d\delta_2(t) = -\frac{\Lambda_2}{\varepsilon}\delta_2(t)dt + [\eta_8 - \eta_4\ \delta_2(t) + (\eta_{10} - \eta_8)p^L(t) + \eta_9\ \delta_1(t)]dt + \frac{g(\varepsilon)}{\sqrt{\varepsilon}}[\xi_3 - \xi_3\ p^L(t) + \xi_4\ \delta_2(t) - \hat{h}(t)\ \delta_2(t)]dw(t) \tag{6-12c}$$

where dw(t) is a standard Brownian motion representing a scaled version of the innovations.

We now turn our attention to the description of two different approximations of the optimal estimator. Both are based on the fact that at the slow time scale the left-right transition behavior is approximately that of a 2-state Markov process [29]. Specifically, consider the time-scaled 4-state

process $\rho(t/\varepsilon)$. Over any interval Δt, this process undergoes numerous (indeed infinitely many as $\varepsilon \to 0$) top-to-bottom transitions. This suggests two important approximations. The first is that if we aggregate away these very fast transitions, we will be left with a two-state Markov process. Specifically, let

$$q(t) = \begin{cases} L & \text{if } \rho(t/\varepsilon) = 1 \text{ or } 2 \\ R & \text{if } \rho(t/\varepsilon) = 3 \text{ or } 4 \end{cases} . \tag{6-13}$$

Then, as shown in [29], $q(t)$ is asymptotically a 2-state Markov process with transition behavior depicted in Fig. 6-2. Here the rate γ_1 represents a weighted average of the left-to-right rates μ_1 and μ_3 in Fig. 6-1, where the weights equal the ergodic probabilities of being in states 1 and 2 when we neglect left-to-right transitions (i.e., the fast top-to-bottom transitions essentially reach equilibrium before a transition from left to right occurs so that the rate of such a transition can be computed by this weighted average). Obviously, there is an analogous interpretation for γ_2.

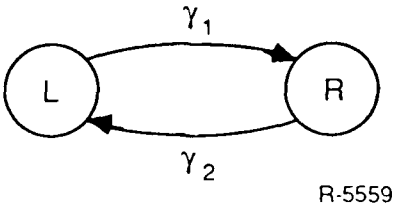


Figure 6-2. The 2-State Aggregate Process

The second approximation is based on the fact that over time interval Δt, the fast top-to-bottom transitions have a similar averaging effect on the observations, i.e., that we can model our observations as

$$dy(t) = g(\varepsilon)\, f[q(t)]dt + dv(t) \tag{6-14}$$

where

$$f(L) = f_L \quad , \quad f(R) = f_R \tag{6-15}$$

again represent averages.

The resulting nonlinear filtering equation in this case is given by

$$dp_L(t) = \varepsilon[\gamma_2 - (\gamma_1 + \gamma_2)p_L(t)]dt + g(\varepsilon)[f_L - \bar{h}(t)]p_L(t)[dy(t) - g(\varepsilon)\bar{h}(t)dt] \tag{6-16}$$

where

$$h(t) = f_R + (f_L - f_R)p_L(t) \quad . \tag{6-17}$$

Comparing Eqs. 6-16 and 6-17 to Eqs. 6-12 and 6-9, we see some important similarities. Indeed, if δ_1 and δ_2 were replaced by zero in Eqs. 6-9 and 6-12, the equations in fact become identical. One might conjecture then that under the right circumstances, Eqs. 6-16 and 6-17 would be an excellent approximation to the optimal estimator of the slow variables. The analysis and simulation results in the following sections show that for measurements with a particular range of signal-to-noise ratios, as characterized by $g(\varepsilon)$, this is in fact the case.

It is instructive to examine the architectural implications of the approximation we have just described. In Fig. 6-3 we have depicted the architecture of the optimal estimator, in which there are both fast and slow variables, while in Fig. 6-4 we have the much simple (pictorially and computationally) approximate estimator. In this estimator we are in essence taking advantage of the slow p_L dynamics to average out the effects of the fast fluctuations in the data. In fact this also suggests a related approximation. In particular, since the approximate estimator has no fast variables, its

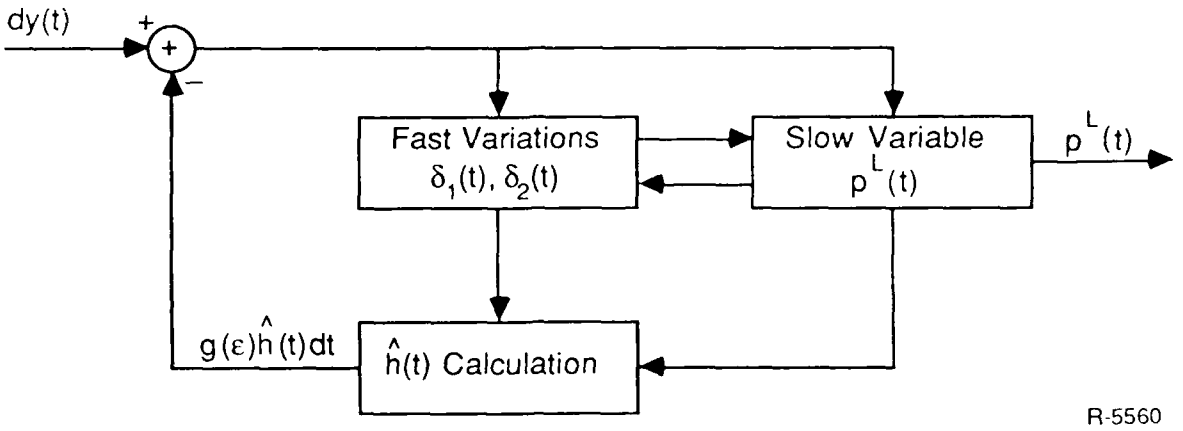


Figure 6-3. Architecture of the Optimal Filter

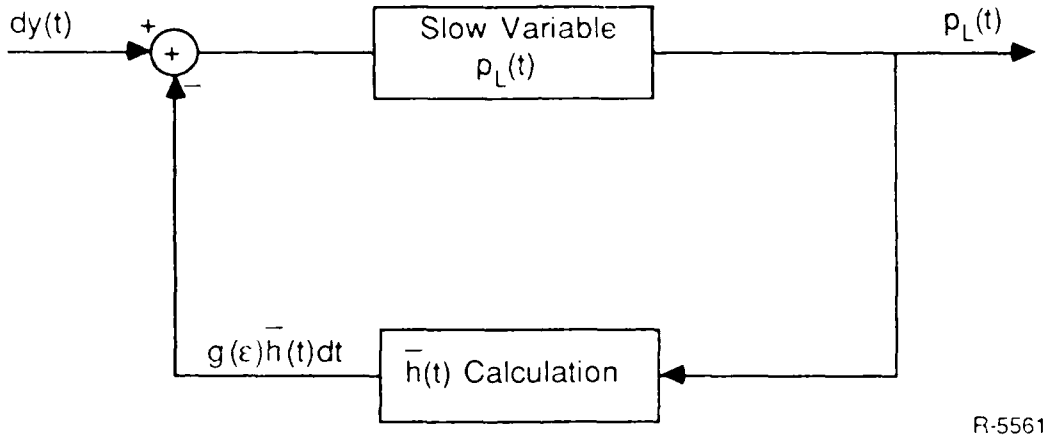


Figure 6-4. Architecture of the First Approximate Filter

digital implementations can use a far larger step size. What this implies is the structure depicted in Fig. 6-5. Here the front-end preprocessor performs an averaging of the observations over a time interval that is long with respect to the fast variables but short with respect to the slow variables. The back-end processor then uses these averaged values in a sampled data approximation of Eq. 6-16.

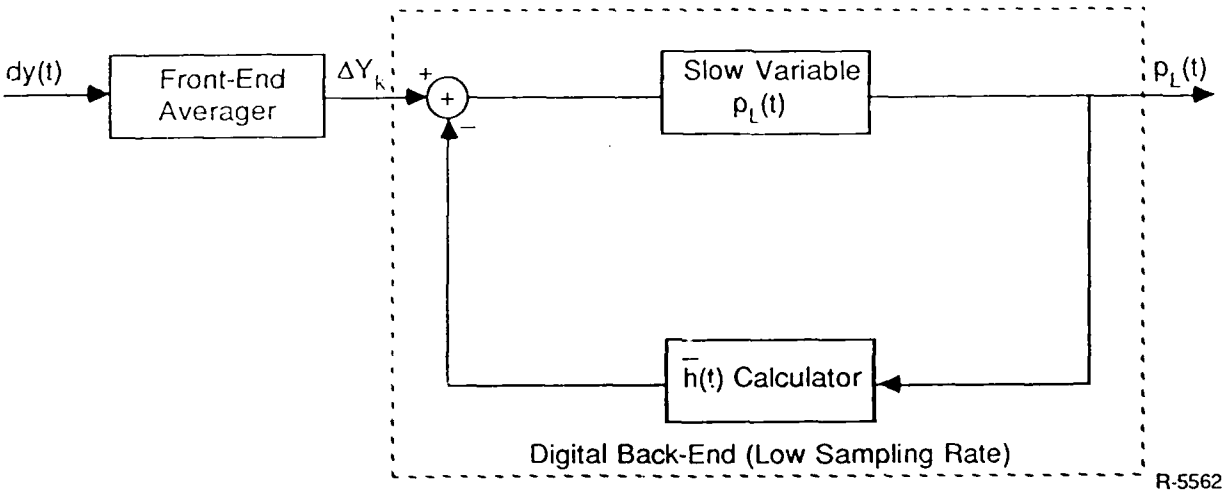


Figure 6-5. A Front-End/Back-End Structure Based on the Approximate Filter in Fig. 6-4

It is also possible to describe a considerably different front-end/back-end structure for an approximate estimator. This structure is motivated explicitly by passive acoustic tracking architectures in which the front end uses a batch of data to estimate slowly-varying quantities -- bearing, frequency -- based on the assumption that these quantities are constant over the time interval of observation. The sequence of estimates produced by the front end then are used as measurements by the back end which attempts to track these slowly varying quantities.

In our present context, a structure of this type would have the form depicted in Fig. 6-6. Here the front-end uses a batch of data for $kT(\varepsilon) \leq t < (k+1)T(\varepsilon)$ to perform an hypothesis test -- is the processor on the right or the left -- based on the assumption that no left-to-right transitions can occur, i.e., on the model depicted in Fig. 6-7. The sufficient statistic for this hypothesis test can be taken to be either P_L, the conditional probability that the process in Fig. 2-7 is on the left, or the likelihood ratio

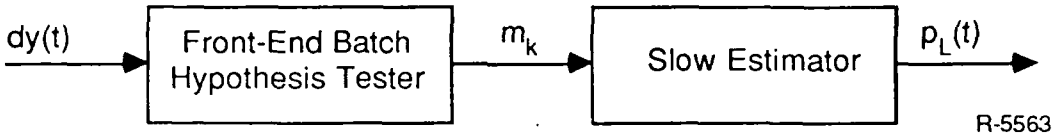


Figure 6-6. Front-End/Back-End Structure with Front-End Hypothesis Testing

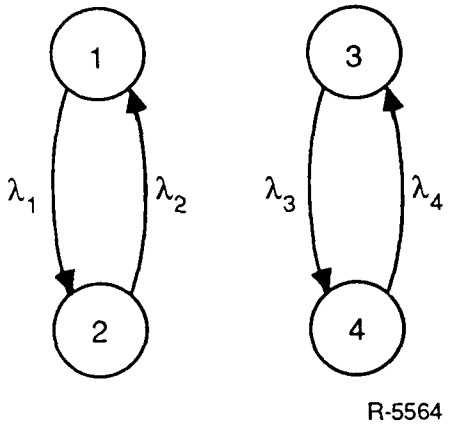


Figure 6-7. Model on Which the Front End of Fig. 6-6 is Based

$$L = \frac{P_L}{1 - P_L} \quad . \tag{6-18}$$

The computation of P_L is a nonlinear filtering problem analogous to the one described previously -- i.e., by getting the $\mu_i = 0$ in Eq. 6-5, letting

$$P_L = (p_1 + p_2) \quad , P_{LT} = p_1 \quad , \quad P_{RT} = p_3 \tag{6-19}$$

and replacing p_2 by $P_L - P_{LT}$ and p_4 by $1 - P_L - P_{RT}$, we obtain

$$dP_L(t) = g(\varepsilon)[\hat{h}_L(t) - \hat{h}(t)]P_L(t)[dy(t) - g(\varepsilon)\,\hat{h}(t)dt] \tag{6-20a}$$

$$\begin{aligned} dP_{LT}(t) = {} & [-\lambda_1 P_{LT}(t) + \lambda_2(P_L(t) - P_{LT}(t))]dt \\ & + g(\varepsilon)[\alpha - \hat{h}(t)]P_{LT}(t)[dy(t) - g(\varepsilon)\,\hat{h}(t)dt] \end{aligned} \tag{6-20b}$$

$$dP_R(t) = [-\lambda_3 P_{RT}(t) + \lambda_4(1 - P_L(t) - P_{RT}(t))]dt + g(\varepsilon)[\alpha - \hat{h}(t)]P_{RT}(t)[dg(t) - g(\varepsilon)\hat{h}(t)dt] \quad (6\text{-}20c)$$

where

$$\hat{h}(t) = \beta + (\alpha - \beta)(P_{LT}(t) + P_{RT}(t)) \quad (6\text{-}21a)$$

$$\hat{h}_L(t) = \beta + (\alpha - \beta)\frac{P_{LT}(t)}{P_L(t)} \quad (6\text{-}21b)$$

Here $\hat{h}_L(t)$ is the expected value of $h(\rho(t))$ given the data and assuming that the process is on the left.

Several alternate forms of these equations are also of potential value both in providing insight and for their potential computational simplicity. First we define the conditional probability of being in state 1 and state 3 given the data and assuming that the process is on the left and right, respectively:

$$Q_{LT} = \frac{P_{LT}}{P_L} \qquad Q_{RT} = \frac{P_{RT}}{1 - P_L} \; . \quad (6\text{-}22)$$

An application of its differential rule [8] then yields

$$dP_L(t) = g(\varepsilon)[\hat{h}_L(t) - h(t)]P_L(t)[dy(t) - g(\varepsilon)\hat{h}(t)dt] \quad (6\text{-}23a)$$

$$dQ_{LT}(t) = [\lambda_2 - (\lambda_1 + \lambda_2)Q_{LT}(t)]dt + g(\varepsilon)(\alpha - \hat{h}_L(t))Q_{LT}(t)(dy(t) - g(\varepsilon)\hat{h}_L(t)dt) \quad (6\text{-}23b)$$

$$dQ_{RT}(t) = [\lambda_4 - (\lambda_3 + \lambda_4)Q_{RT}(t)]dt + g(\varepsilon)(\alpha - \hat{h}_R(t))Q_{RT}(t)(dy(t) - g(\varepsilon)\hat{h}_R(t)dt) \quad (6\text{-}23c)$$

where

$$\hat{h}_L(t) = \beta + (\alpha - \beta)\, Q_{LT}(t) \tag{6-24a}$$

$$\hat{h}_R(t) = \beta + (\alpha - \beta)\, Q_{RT}(t) \tag{6-24b}$$

$$\hat{h}(t) = \hat{h}_L(t)\, P_L(t) + \hat{h}_R(t)\, (1 - P_L(t)) \quad . \tag{6-24c}$$

Note the decoupled structure of these equations as depicted in Fig. 6-8, where we have used the fact that

$$d\nu(t) = dy(t) - g(\varepsilon)\, \hat{h}(t)dt = P_L(t)\, d\nu_L(t) + (1 - P_L(t)d\nu_R(t) \tag{6-25a}$$

where

$$d\nu_R(t) = dy(t) - g(\varepsilon)\, \hat{h}_R(t)dt \quad , \quad d\nu_L(t) = dy(t) - g(\varepsilon)\, \hat{h}_L(t)dt \tag{6-25b}$$

Note also that unlike the exact nonlinear filter (Eq. 6-5), these equations do not involve variables at 2 time scales and thus can be integrated more efficiently.

Yet another useful form can be obtained using the likelihood ratio (Eq. 6-18):

$$dL(t) = g(\varepsilon)\, (\hat{h}_L(t) - \hat{h}_R(t))\, L(t)d\nu_R(t) \tag{6-26a}$$

$$dQ_{LT}(t) = [\lambda_2 - (\lambda_1 + \lambda_2)Q_{LT}(t)]dt + g(\varepsilon)(\alpha - \hat{h}_L(t))Q_{LT}(t)\, d\nu_L(t) \tag{6-26b}$$

$$dQ_{RT}(t) = [\lambda_4 - (\lambda_3 + \lambda_4)Q_{RT}(t)]dt + g(\varepsilon)(\alpha - \hat{h}_R(t))Q_{RT}\, d\nu_R(t) \tag{6-26c}$$

Note that these quations are somewhat simpler than the preceding set since we do not need to compute $d\nu(t)$. Taking this one step further, if we let

$$\ell(t) = \log L(t) \tag{6-27}$$

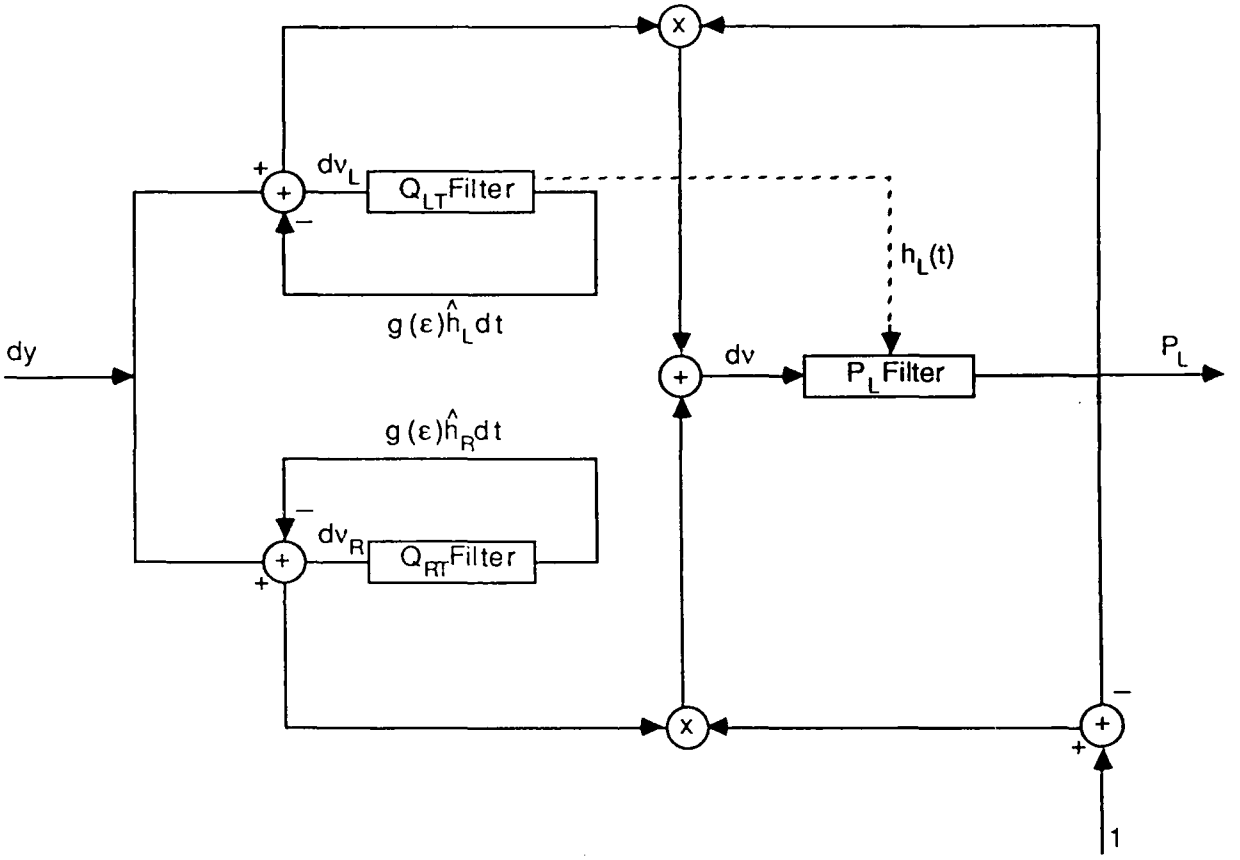


Figure 6-8. The Architecture of the Front-End in Fig. 6-6

then

$$d\ell(t) = -\frac{1}{2} g^2(\varepsilon)[\hat{h}_L(t) - \hat{h}_R(t)]^2 dt + g(\varepsilon)(\hat{h}_L(t) - \hat{h}_R(t)]d\nu_R(t) \qquad (6\text{-}28)$$

which is just a simple accumulation:

$$\ell((k+1)T(\varepsilon)) = \ell(kT(\varepsilon)) - \frac{1}{2} g^2(\varepsilon) \int_{kT(\varepsilon)}^{(k+1)T(\varepsilon)} [\hat{h}_L(t) - \hat{h}_R(t)]^2 dt + g(\varepsilon) \int_{kT(\varepsilon)}^{(k+1)T(\varepsilon)} [\hat{h}_L(t) - \hat{h}_R(t)]d\nu_R(t) \qquad (6\text{-}29)$$

Let us comment on the choice of initial conditions at the start of each interval. A natural set of choices are

$$Q_{LT} = \frac{\lambda_2}{\lambda_1 + \lambda_2} \quad , \quad Q_{RT} = \frac{\lambda_4}{\lambda_3 + \lambda_4} \qquad (6\text{-}30a)$$

and the following equivalent conditions

$$P_L = \frac{1}{2} \quad , \quad L = 1 \quad , \quad \ell = 0 \quad . \qquad (6\text{-}30b)$$

Equation 6-30a corresponds to assuming that the fast process is in equilibrium when we begin an observation interval. Equation 6-30b corresponds to no prior information, i.e., no feedback is provided from the back end in Fig. 6-6 to front end.

Finally, at the end of each interval, we perform a threshold test. In terms of ℓ, this is

$$\ell[(k+1)T(\varepsilon)] \begin{cases} > \\ \\ < \end{cases} \begin{matrix} m_{k+1} = L \\ \\ 0 \\ \\ m_{k+1} + r \end{matrix} \qquad (6\text{-}31)$$

where "L" and "R" are the two values of the observational input to the back end in Fig. 6-6. The back end is then simply a sampled data estimator of the 2-state process in Fig. 6-2, with measurements provided every $T(\varepsilon)$ time units. We can characterized the performance of the front end in terms of 2 quantities

$$\phi_{LL}(\varepsilon) = \text{Prob}[m_k = L \mid \text{process on the left}] \qquad (6\text{-}32a)$$

$$\phi_{LR}(\varepsilon) = \text{Prob}[m_k = L \mid \text{process on the right}] \qquad (6\text{-}32b)$$

We explicitly write these as functions of ε to indicate that the performance of the test (Eq. 6-31) depends upon the size of the time interval. Note, however, that these performance measures only make sense for $T(\varepsilon)$ small with respect to $1/\varepsilon$ so that the process actually is very likely to stay on the left or on the right over the entire interval.

Given the quantities in Eq. 6-32, the optimal back end can be described as follows. Let

$$p_L(k|j) = \text{Prob}[\text{process on left at time } kT(\varepsilon) \mid m_s \ , \ s < j] \qquad (6\text{-}33)$$

then

$$p_L(k+1|k) = e^{-(\gamma_1+\gamma_2)T(\varepsilon)} p_L(k|k) + \frac{\gamma_2}{\gamma_1 + \gamma_2}\left(1 - e^{-(\gamma_1+\gamma_2)T(\varepsilon)}\right) \qquad (6\text{-}34)$$

and if $m_{k+1} = L$

$$p_L(k+1|k+1) = \frac{\phi_{LL}(\varepsilon)p_L(k+1|k)}{\phi_{LL}(\varepsilon)p_L(k+1|k) + \phi_{LR}(\varepsilon)\ [1 - p_L(k+1|k)]} \tag{6-35a}$$

while if $m_{k+1} = R$

$$p_L(k+1|k+1) = \frac{[1 - \phi_{LL}(\varepsilon)]p_L(k+1|k)}{[1 - \phi_{LL}(\varepsilon)]p_L(k+1|k) + [1 - \phi_{LR}(\varepsilon)][1 - p_L(k+1|k)]} \tag{6-35b}$$

The evaluation of $\phi_{LL}(\varepsilon)$ and $\phi_{LR}(\varepsilon)$ are therefore essential both to define the update step (Eq. 6-35) and to evaluate its overall performance and asymptotic properties. A key question here involves the relationships among the time scale separation controlled by ε, the rate at which information is obtained controlled by $g(\varepsilon)$, and the data collection interval $T(\varepsilon)$.

We note that it is also possible to define a slightly different front-end/back-end structure in which we eliminate the hard decision (Eq. 6-31) in the front end and instead take as input to the back end the likelihood ratio

$$L_k = L(kT(\varepsilon)) \quad . \tag{6-36}$$

Note that for $T(\varepsilon)$ small with respect to $1/\varepsilon$ and with an initial condition of 1 on L at the start of an interval, L_k is approximately equal to

$$\frac{\Pr\ [\text{data over } [(k-1)T(\varepsilon),kT(\varepsilon)) \mid \text{process on the left}]}{\Pr\ [\text{data over } [(k-1)T(\varepsilon),kT(\varepsilon)) \mid \text{process on the right}]}$$

for the true process. This suggests the following back end algorithm. The prediction step is still given by Eq. 6-34. The update step, however, is given by

$$p_L(k+1|k+1) = \frac{L_{k+1}\ p_L(k+1|k)}{L_{k+1}\ p_L(k+1|k) + (1 - p_L(k+1|k))} \quad . \tag{6-37}$$

Finally, we note that we can combine our two types of approximations to obtain a far simpler one. Specifically, suppose we model our observations as in Eq. 6-14 by assuming an averaging effect. Writing these in the original time scale

$$dy(t) = g(\varepsilon)\, f(q(t))dt + dv(t) \quad . \tag{6-38}$$

Then if our front end assumes no left-to-right transitions, we obtain

$$dp_L = g(\varepsilon)[f_L - \bar{h}(t)]\, p_L(t)[dy(t) - g(\varepsilon)\, \bar{h}(t)dt] \tag{6-39}$$

$$\bar{h}(t) = f_R + (f_L - f_R)\, p_L(t) \quad . \tag{6-40}$$

If we define

$$L(t) = \frac{p_L(t)}{1 - p_L(t)} \tag{6-41}$$

we obtain

$$dL(t) = g(\varepsilon)\, (f_L - f_R)\, L(t)[dy(t) - g(\varepsilon) f_R dt] \tag{6-42}$$

or if

$$\ell(t) = \log L(t) \tag{6-43}$$

then

$$d\ell(t) = \frac{1}{2}\, g^2(\varepsilon)\, [f_L - f_R]dt^2 + g(\varepsilon)\, (f_L - f_R)\, [dy(t) - g(\varepsilon) f_R dt] \tag{6-44}$$

i.e.,

$$d\ell(t) = \frac{1}{2}\, g^2(\varepsilon)\, [f_R^2 - f_L^2] + g(\varepsilon)\, (f_L - f_R)dy(t) \quad . \tag{6-45}$$

This leads to the following algorithm: use Eq. 6-34 for the prediction step and Eq. 6-37 for the measurement update, with

$$L_k = e^{\ell_k} \tag{6-46}$$

$$\ell_k = \frac{1}{2} g^2(\varepsilon) [f_R^2 - f_L^2]T(\varepsilon) + g(\varepsilon) [f_L - f_R] \int_{kT(\varepsilon)}^{(k+1)T(\varepsilon)} dy(\tau) \quad . \tag{6-47}$$

Note that the basic processing step in the front end is exactly a simple averaging of the measurements.

6.2.2 Perfect Measurements with Small Differences in Rates

In this case, since we directly observe whether the process is on the top or the bottom, the exact filtering equations involve only $p_L(t)$, the probability of being on the left. To facilitate our development, we introduce a process indicating the top-bottom status of the processor

$$x(t) = \begin{cases} 1 \ , & \rho(t) = 1 \text{ or } 3 \\ 0 \ , & \rho(t) = 2 \text{ or } 4 \ . \end{cases} \tag{6-48}$$

Then, using the analysis in Appendix B we have the following description of the evolution of $p_L(t)$:

While $x(t) = 1$

$$\dot{p}_L(t) = \varepsilon\mu_2 - \varepsilon(\mu_1 + \mu_2)\, p_L(t) + (\lambda_3 - \lambda_1)\, p_L(t)[1 - p_L(t)] \tag{6-49a}$$

While $x(t) = 0$

$$\dot{p}_L(t) = \varepsilon\mu_4 - \varepsilon(\mu_3 + \mu_4)\, p_L(t) + (\lambda_4 - \lambda_2)\, p_L(t)[1 - p_L(t)] \tag{6-49b}$$

When $x(t-) = 1$ and $x(t) = 0$

$$p_L(t) = \frac{\lambda_1 \, p_L(t-)}{(\lambda_1 - \lambda_3) \, p_L(t-) + \lambda_3} \tag{6-49c}$$

When $x(t-) = 0$ and $x(t) = 1$

$$p_L(t) = \frac{\lambda_2 \, p_L(t-)}{(\lambda_2 - \lambda_4) \, p_L(t-) + \lambda_4} \, . \tag{6-49d}$$

These can be combined into the following stochastic differential equation where we also use Eq. 6-3:

$$\begin{aligned} dp_L(t) = {} & [\varepsilon\mu_2 - \varepsilon(\mu_1 + \mu_2) \, p_L(t) = \alpha g(\varepsilon) \, p_L(t) \, (1 - p_L(t))]x(t)dt \\ & + [\varepsilon\mu_4 - \varepsilon(\mu_3 + \mu_4) \, p_L(t) + \beta g(\varepsilon) \, p_L(t) \, (1 - p_L(t))][1 - x(t)]dt \\ & + g(\varepsilon)p_L(t)[1-p_L(t)] \left[\frac{\alpha x(t)}{\lambda_1+\alpha g(\varepsilon)[1-p_L(t)]} - \frac{\beta[1-x(t)]}{\lambda_2+\beta g(\varepsilon)[1-p_L(t)]} \right] dx(t) \end{aligned} \tag{6-50}$$

Two alternate forms of these equations provide additional insight. First note that

$$\begin{aligned} E[dx(t) | x(\tau), \, \tau<t] = {} & -x(t)[\lambda_1 \, p_L(t) + \lambda_3(1 - p_L(t))]dt \\ & + [1 - x(t)] \, [\lambda_2 \, p_L(t) + \lambda_4(1 - p_L(t))]dt \\ = {} & -x(t)[\lambda_1 + \alpha g(\varepsilon) \, (1 - p_L(t))] \\ & + [1 - x(t)] \, [\lambda_2 + \beta g(\varepsilon) \, (1 - p_L(t))] \\ \triangleq {} & \hat{h}(t)dt \end{aligned} \tag{6-51}$$

Using Eqs. 6-3 and 6-51 we can rewrite Eq. 6-50 as

$$dp_L(t) = \varepsilon[\mu_2 - (\mu_1 + \mu_2)p_L(t)]x(t)dt + \varepsilon[\mu_4 - (\mu_3 + \mu_4)p_L(t)][1 - x(t)]dt$$

$$+ g(\varepsilon)p_L(t)[1-p_L(t)]\left[\frac{\alpha x(t)}{\lambda_1+\alpha g(\varepsilon)[1-p_L(t)]} - \frac{\beta[1-x(t)]}{\lambda_2+\beta g(\varepsilon)[1+p_L(t)]}\right](dx(t)-\hat{h}(t)dt) \tag{6-52}$$

A second important form involves two derived quantities

$$R(t) = \text{Total residence time of } \rho(\tau)\ 0<\tau<t \text{ on the top (i.e., in states 1 or 3) .} \tag{6-53}$$

$$K(t) = \text{Total number of changes in } x(\tau) \text{ over } 0<\tau<t \ .$$

Then

$$dR(t) = x(t)dt$$
$$dK(t) = |dx(t)| = -x(t)dx(t) + [1-x(t)]dx(t) = [1-2x(t)]dx(t) \ . \tag{6-54}$$

Note several equalities

$$x^2(t) = x(t) \ , \ [1-x(t)]^2 = 1-x(t) \ , \ x(t)[1-x(t)] = 0$$
$$[1-2x(t)]^2 = 1 \ , \ x(t)[1-2x(t)] = -x(t) \ , \ [1-x(t)][1-2x(t)] = 1-x(t) \tag{6-55}$$

Using Eqs. 6-50, 6-54, and 6-55 we obtain

$$dp_L(t) = [\varepsilon\mu_4 - \varepsilon(\mu_3 + \mu_4)p_L(t) + \beta g(\varepsilon)p_L(t)\ (1 - p_L(t))]dt$$
$$+ [\varepsilon(\mu_2 - \mu_4) - \varepsilon(\mu_1 + \mu_2 - \mu_3 - \mu_4)\ p_L(t)$$
$$+ (\alpha-\beta)\ g(\varepsilon)\ p_L(t)\ (1 - p_L(t))]dR(t)$$
$$- g(\varepsilon)p_L(t)[1-p_L(t)]\left[\frac{\alpha x(t)}{\lambda_1+\alpha g(\varepsilon)[1-p_L(t)]} + \frac{\beta[1-x(t)]}{\lambda_2+\beta g(\varepsilon)[1-p_L(t)]}\right]dK(t) \tag{6-56}$$

Equation 6-56 suggests one straightforward front-end/back-end decomposition. Specifically, because of the slow dynamics in Eq. 2-56 one can imagine a system of the form depicted in Fig. 6-9. Here the front end performs a simple accumulation, computing ΔR and ΔK over an interval of time that is long with respect to the fast dynamics and short with respect to the slow dynamics. These quantities are then fed into a sampled version of Eq. 6-56.

One can also envision two other, more sophisticated front-end/back-end approximations to Eq. 6-56, each of which is analogous to one of the forms described in the preceding subsection. In the first of these, depicted in Fig. 6-10, we perform a front-end hypothesis test whose results are fed into a back-end slow estimator. Comparing Figs. 6-6 and 6-10, we see that there is a difference in that the slow estimator uses both the hypothesis test results and the raw data $x(t)$. To understand this, it is important to realize that $x(t)$ provides us with two types of information: the indirect information about whether $\rho(t)$ is on the left or right that is embedded in the switching behavior of $x(t)$ and the <u>direct</u> information concerning which left-to-right rates -- μ_1 and μ_2 or μ_3 and μ_4 -- are in effect. Indeed, consider the two-state process depicted in Fig. 6-11. The evolution of the unconditional probability of being on the left for this process is given by

$$\begin{aligned} dp_L(t) &= -[\varepsilon\mu_1 x(t) + \varepsilon\mu_3(1-x(t))]p_L(t) + [\varepsilon\mu_2 x(t) + \varepsilon\mu_4(1-x(t))][1-p_L(t)]dt \\ &= [\varepsilon\mu_4 - \varepsilon(\mu_3+\mu_4)p_L(t)]dt + [\varepsilon(\mu_2 - \mu_4) - \varepsilon(\mu_1+\mu_2-\mu_3-\mu_4)p_L(t)]dR(t) \end{aligned} \tag{6-57}$$

Compare this to Eq. 6-56. Note that the remaining terms in Eq. 6-56 -- i.e., those not captured in Eq. 6-57, yield the equation for the conditional probability of being on the left given the data and assuming no left-to-right transitions occur:

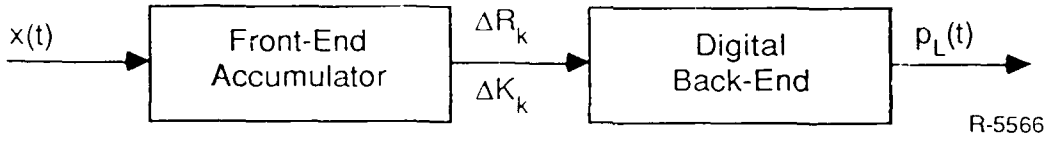


Figure 6-9. Front-End/Back-End Structure Arising from Slow Integration of Optimal Estimator

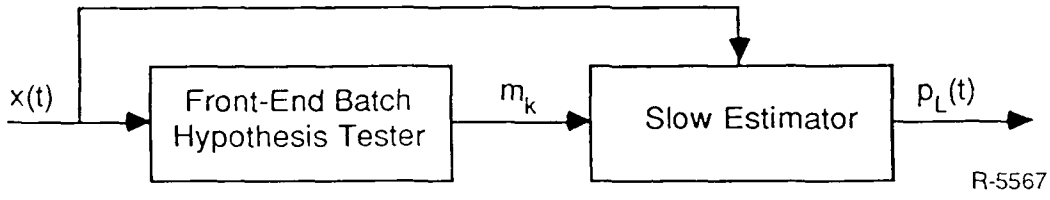


Figure 6-10. Hypothesis-Testing Front End with Slow Back-End Estimator

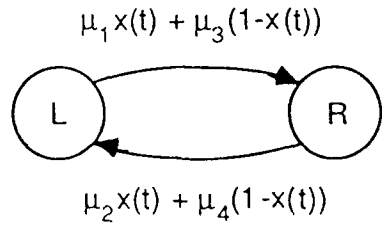
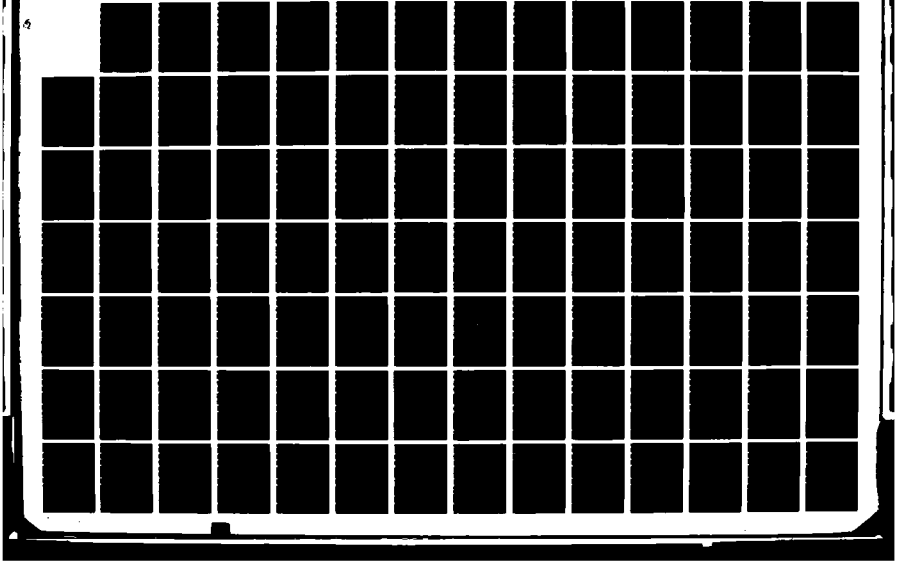
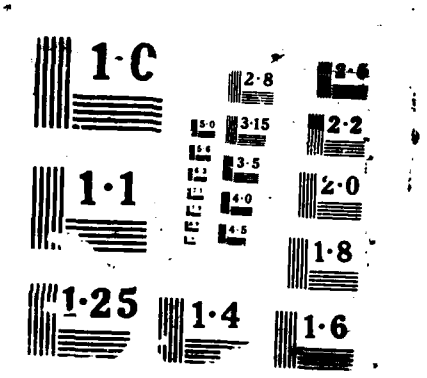


Figure 6-11. The Left-Right Transition Rates

AD-A188 424 PASSIVE ACOUSTIC TRACKING AND FILTERING PROBLEMS WITH TIME SCALES(U) ALPHATECH INC BURLINGTON MA A CARONICOLI ET AL NOV 87 TR-352 N00014-85-C-0349 2/3

UNCLASSIFIED F/G 17/1 NL




1·C
2·8
2·6
3·15
2·2
3·5
1·1
4·0
2·0
4·5
1·8
1·25
1·4
1·6

$$dp_L(t) = \beta g(\varepsilon) p_L(t)[1-p_L(t)]dt + (\alpha-\beta) g(\varepsilon) p_L(t)[1-p_L(t)]dR(t)$$

$$- g(\varepsilon) p_L(t)[1-p_L(t)\left[\frac{\alpha x(t)}{\lambda_1+\alpha g(\varepsilon)[1-p_L(t)]} + \frac{\beta[1-x(t)]}{\lambda_2+\beta g(\varepsilon)[1-p_L(t)]}\right] dK(t) \quad . \tag{6-58}$$

Equation 6-58 can in fact be solved in closed form, as the sufficient statistics in this case are essentially R(t) and K(t). Specifically, let

$$\begin{aligned} D(t) &= \text{Number of top-to-bottom transitions up to time } t \\ U(t) &= \text{Number of bottom-to-top transitions up to time } t \quad . \end{aligned} \tag{6-59}$$

Note that

$$K(t) = D(t) + U(t) \tag{6-60}$$

and if $x(0) = 1$

$$D(t) = \left\lceil \frac{K(t)}{2} \right\rceil \quad , \quad U(t) = \left\lfloor \frac{K(t)}{2} \right\rfloor \tag{6-61a}$$

while if $x(0) = 0$

$$D(t) = \left\lfloor \frac{K(t)}{2} \right\rfloor \quad , \quad U(t) = \left\lceil \frac{K(t)}{2} \right\rceil \tag{6-61b}$$

where

$$\begin{aligned} \lceil x \rceil &= \text{smallest integer} \geq x \\ \lfloor x \rfloor &= \text{largest integer} \leq x \quad . \end{aligned} \tag{6-62}$$

Define

$$L(t) = e^{-\beta g(\varepsilon) t} \, e^{(\beta-\alpha) g(\varepsilon) R(t)} \left(1 + \frac{\alpha g(\varepsilon)}{\lambda_1}\right)^{D(t)} \left(1 + \frac{\beta g(\varepsilon)}{\lambda_2}\right)^{U(t)} \quad . \tag{6-63}$$

Then L(t) is the likelihood ratio

$$L(t) = \frac{\Pr(x(\tau),\ \tau<t \mid \rho(t) \text{ on the right})}{\Pr(x(\tau),\ \tau<t \mid \rho(t) \text{ on the left})} \tag{6-64}$$

assuming that no left-to-right transitions can take place. In this case, it is easy to see that

$$p_L(t) = \frac{p_L(0)}{p_L(0) + [1-p_L(0)]L(t)} \ . \tag{6-65}$$

It is a straightforward exercise to verify that Eq. 6-55 satisfies Eq. 6-58.

We can now describe precisely the front-end/back-end structure depicted in Fig. 6-10. Again, let $(k-1)\,T(\varepsilon) < t < kT(\varepsilon)$ denote the k-th batch of data for the front-end processor, and let

$$R_k = R(kT(\varepsilon)) - R[(k-1)T(\varepsilon)] \tag{6-66a}$$

$$D_k = D(kT(\varepsilon)) - D[(k-1)T(\varepsilon)] \tag{6-66b}$$

$$U_k = U(kT(\varepsilon)) - U[(k-1)T(\varepsilon)] \tag{6-66c}$$

$$K_k = K(kT(\varepsilon)) - K[(k-1)T(\varepsilon)] \tag{6-66d}$$

$$L_k = \text{likelihood ratio based on the k-th batch of data} \tag{6-66e}$$

i.e.,

$$L_k = e^{-\beta g(\varepsilon)T(\varepsilon)} e^{(\beta-\alpha)g(\varepsilon)R_k} \left(1 + \frac{\alpha g(\varepsilon)}{\lambda_1}\right)^{D_k} \left(1 + \frac{\beta g(\varepsilon)}{\lambda_2}\right)^{U_k} \ . \tag{6-67}$$

The front end then produces an output based on the decision rule

$$L_k \begin{cases} > \\ < \end{cases} 1 \quad \begin{matrix} m_k = R \\ m_k = L \end{matrix} \tag{6-68}$$

Also let

$$p_L(k|j) = \text{Prob}[\text{process on left at time } kT(\varepsilon) \,|m_s,\ s\leq j] \quad . \tag{6-69}$$

The prediction step of the algorithm consists of integrating

$$\begin{aligned} \dot{p}_L = &- [\varepsilon\mu_1 x(t) + \varepsilon\mu_3(1-x(t))]p_L \\ &+ \varepsilon[\mu_2 x(t)+\varepsilon(1-x(t))](1-p_L) \end{aligned} \tag{6-70}$$

from $t = kT(\varepsilon)$ to $t = (k+1)T(\varepsilon)$ using as initial condition $p_L(k|k)$. The update step is then

$$p_L(k+1|k+1) = \frac{\phi_{LL}(\varepsilon)\ p_L(k+1|k)}{\phi_{LL}(\varepsilon)\ p_L(k+1|k) + \phi_{LR}[1-p_L(k+1|k)]} \tag{6-71a}$$

if $m_k = L$, and

$$p_L(k+1|k+1) = \frac{[1-\phi_{LL}(\varepsilon)]\ p_L(k+1|k)}{[1-\phi_{LL}(\varepsilon)]p_L(k+1|k) + [1-\phi_{LR}(\varepsilon)][1-p_L(k+1|k)]} \tag{6-71b}$$

if $m_k = R$, where ϕ_{LL} and ϕ_{LR} are defined in Eq. 6-32. We note that in this case it is possible to carry out some of the analysis associated with these quantities, and this is done in Appendix C.

Note that in general the exact computation of the prediction equation (Eq. 6-70) involves the complete x(t) sample path -- essentially we are switching between two sets of rates, i.e.,

$$\begin{bmatrix} \dot{p}_L(t) \\ \dot{p}_R(t) \end{bmatrix} = \varepsilon[A_1\, x(t) + A_2[1-x(t)]] \begin{bmatrix} p_L(t) \\ p_R(t) \end{bmatrix} \tag{6-72}$$

where

$$A_1 = \begin{bmatrix} -\mu_1 & \mu_2 \\ \mu_1 & -\mu_2 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -\mu_3 & \mu_4 \\ \mu_3 & -\mu_4 \end{bmatrix}. \tag{6-73}$$

If

$$\mu_1/\mu_2 = \mu_3/\mu_4 \tag{6-74}$$

i.e., if the steady-state probabilities associated with A_1 and A_2 are the same, then A_1 and A_2 commute and the solution of Eq. 6-72 involves only the integral of x(t), namely R(t). If Eq. 6-74 does not hold, this isn't true exactly, although for $\varepsilon T(\varepsilon)$ small, the difference between the solution to Eq. 6-72 at $t = (k+1)T(\varepsilon)$ from initial condition $p_L(k|k)$ at $t = kT(\varepsilon)$ and the approximation

$$[1,0]\, \exp\{\varepsilon A_1\, R_{k+1} + \varepsilon A_2[T(\varepsilon) - R_{k+1}]\} \begin{bmatrix} p_L(k|k) \\ 1-p_L(k|k) \end{bmatrix} \tag{6-75}$$

is $O(\varepsilon^2\, T^2(\varepsilon))$. If this is ignored, we can use Eq. 6-75 which in expanded form is as follows:

$$\begin{aligned} p_L(k+1|k) = {} & e^{-\varepsilon[\psi_1\, T(\varepsilon) + \psi_2\, R_{k+1}]} p_L(k|k) \\ & + \frac{[\mu_4\, T(\varepsilon) + (\mu_1-\mu_3)R_{k+1}][1 - e^{-\varepsilon[\psi_1\, T(\varepsilon) + \psi_2\, R_{k+1}]}]}{\varepsilon\psi_1\, T(\varepsilon) + \varepsilon\psi_2\, R_{k+1}} \end{aligned} \tag{6-76}$$

where

$$\psi_1 = \mu_3 + \mu_4 \quad , \quad \psi_2 = \mu_1 + \mu_2 - \mu_3 - \mu_4 \quad . \tag{6-77}$$

Note in this case that only R_{k+1}, and not the entire x(t) sample path, is used by the slow estimator.

Finally, as in the preceding subsection, we can replace the hard decision rule (Eq. 6-68) by a Bayesian update. In this case, the front end again calculates R_k, D_k, U_k and L_k as in Eq. 6-67. The back end then performs a prediction step using Eq. 6-70 or the simpler (and exact if Eq. 6-74 holds and otherwise approximate) version (Eq. 6-76) and a measurement update step of the following form:

$$p_L(k+1|k+1) = \frac{p_L(k+1|k)}{p_L(k+1|k) + [1 - p_L(k+1|k)]L_{k+1}} \quad . \tag{6-78}$$

SECTION 7

ASYMPTOTIC ANALYSIS

In this section we present some results on the asymptotic analysis of the white noise filtering problem described in subsections 6.1.1 and 6.2.1. To facilitate the analysis, we repeat the key equations here in somewhat more compact form. In Eqs. 6-8 and 6-9 we presented the exact optimal filtering equations in a particular set of coordinate, while in Eqs. 6-16 and 6-17 we presented the corresponding equation for a particular aggregate approximation. By eliminating dy(t) in Eq. 6-16 through the use of Eq. 6-9a, we obtain the following:

$$dp^L(t) = \varepsilon A\big(p^L(t),\delta_1(t),\delta_2(t)\big)dt + g(\varepsilon)\, B\big(p^L(t),\delta_1(t),\delta_2(t)\big)d\nu(t) \tag{7-1a}$$

$$dp_L(t) = \varepsilon A\big(p^L(t),0,0\big)dt + g(\varepsilon)\, B\big(p^L(t),0,0\big)d\nu(t) + g^2(\varepsilon)\, C\big(p_L(t)p^L(t),\delta_1(t),\delta_2(t)\big)dt \tag{7-1b}$$

$$d\delta_1(t) = -\Lambda_1\, \delta_1(t)dt + \varepsilon D\big(p^L(t),\delta_1(t),\delta_2(t)\big)dt + g(\varepsilon)\, F\big(p^L(t),\delta_1(t),\delta_2(t)\big)d\nu(t) \tag{7-1c}$$

$$d\delta_2(t) = -\Lambda_2\, \delta_2(t)dt + \varepsilon G\big(p^L(t),\delta_1(t),\delta_2(t)\big)dt + g(\varepsilon)\, J\big(p^L(t),\delta_1(t),\delta_2(t)\big)d\nu(t)\ . \tag{7-1d}$$

where $d\nu(t)$ in Eq. 6-9a is a standard Brownian notion and where

$$A(p^L,\delta_1,\delta_2) = \gamma_2 - (\gamma_1 + \gamma_2)p^L - \eta_1\,\delta_1 + \eta_2\,\delta_2 \tag{7-2a}$$

$$B(p^L,\delta_1,\delta_2) = [f_L - H(p^L,\delta_1,\delta_2)]p^L + \Delta\delta_1 \tag{7-2b}$$

$$H(p^L,\delta_1,\delta_2) = f_R + (f_L - f_R)p^L + \Delta[\delta_1 + \delta_2] \tag{7-2c}$$

$$C(p_L,p^L,\delta_1,\delta_2) = [f_L - H(p_L,0,0)]p_L\,[H(p_L,0,0) - H(p^L,\delta_1,\delta_2)] \tag{7-2d}$$

$$D(p^L,\delta_1,\delta_2) = \eta_5 - \eta_3\,\delta_1 + (\eta_4 - \eta_5)p^L + \eta_6\,\delta_2 \tag{7-2e}$$

$$F(p^L,\delta_1,\delta_2) = \xi_1\,p^L + \xi_2\,\delta_1 - H(p^L,\delta_1,\delta_2)\delta_1 \tag{7-2f}$$

$$G(p^L,\delta_1,\delta_2) = \eta_8 - \eta_4\,\delta_2 + (\eta_{10} - \eta_8)p^L + \eta_9\,\delta_1 \tag{7-2g}$$

$$J(p^L,\delta_1,\delta_2) = \xi_3 - \xi_3\,p^L + \delta_4\,\delta_2 - H(p^L,\delta_1,\delta_2)\delta_2 \tag{7-2h}$$

Note that the aggregate approximate probability p_L is now coupled to the exact quantities, as we want to write all equations in a form driven by the true innovations.

What we show in this section is that the approximate filter (Eqs. 6-16 and 6-17) is a good one, in that

$$q(t) = p^L(t) - p_L(t) \tag{7-3}$$

is small compared to the amount of information contained in p^L or p_L. More precisely, let

$$\mu(t) = p^L(t) - \frac{\gamma_2}{\gamma_1 + \gamma_2} = p^L(t) - \psi \quad . \tag{7-4}$$

Then $\mu(t)$ is precisely the deviation of $p^L(t)$ from its steady-state value <u>if no measurements were available</u>. We essentially show that $q(t)$ is quite small

in relation to $\mu(t)$. In preparation for a precise statement of this type, we transform variables in Eq. 7-1 using Eqs. 7-3 and 7-4 and also change to the slow time scale. The result is the following set of equations on which our analysis will focus:

$$d\mu(t) = -\Gamma\,\mu(t)dt + K\big(\delta_1(t),\delta_2(t)\big)dt + \frac{g(\varepsilon)}{\sqrt{\varepsilon}}\,N\big(\mu(t),\delta_1(t),\delta_2(t)\big)dw(t) \tag{7-5a}$$

$$d\delta_1(t) = -\frac{\Lambda_1}{\varepsilon}\,\delta_1(t)dt + P\big(\mu(t),\delta_1(t),\delta_2(t)dt\big) + \frac{g(\varepsilon)}{\sqrt{\varepsilon}}\,Q\big(\mu(t),\delta_1(t),\delta_2(t)\big)dw(t) \tag{7-5b}$$

$$d\delta_2(t) = -\frac{\Lambda_2}{\varepsilon}\,\delta_2(t)dt + R\big(\mu(t),\delta_1(t),\delta_2(t)dt\big) + \frac{g(\varepsilon)}{\sqrt{\varepsilon}}\,S\big(\mu(t),\delta_1(t),\delta_2(t)\big)dw(t) \tag{7-5c}$$

$$dq(t) = -\Gamma\,q(t)dt + K\big(\delta_1(t),\delta_2(t)\big)dt + \frac{g^2(\varepsilon)}{\varepsilon}\,M\big(q(t),\delta_1(t),\delta_2(t)\big)dt + \frac{g(\varepsilon)}{\sqrt{\varepsilon}}\,U\big(q(t),\mu(t),\delta_1(t),\delta_2(t)\big)dw(t) \tag{7-5d}$$

where $w(t)$ is a standard Brownian motion, and

$$\Gamma = \gamma_1 + \gamma_2 \tag{7-6a}$$

$$K(\delta_1,\delta_2) = -\eta_1\,\delta_1 + \eta_2\,\delta_2 \tag{7-6b}$$

$$M(q,\delta_1,\delta_2) = (f_L - f_R)q + \Delta(\delta_1 + \delta_2) \tag{7-6c}$$

$$N(\mu,\delta_1,\delta_2) = [(f_L - f_R)(1-\psi-\mu) - \Delta(\delta_1 + \delta_2)][\mu+\psi] + \Delta\delta_1 \tag{7-6d}$$

$$P(\mu,\delta_1,\delta_2) = \eta_5 - \eta_3\,\delta_1 + (\eta_4 - \eta_5)(\mu+\psi) + \eta_6\,\delta_2 \tag{7-6e}$$

$$Q(\mu,\delta_1,\delta_2) = \xi_1(\mu+\psi) + \xi_2\,\delta_1 - [f_R + (f_L - f_R)(\mu+\psi) + \Delta(\delta_1+\delta_2)]\delta_1 \tag{7-6f}$$

$$R(\mu,\delta_1,\delta_2) = \eta_8 - \eta_7\,\delta_2 + (\eta_{10} - \eta_8)(\mu+\psi) + \eta_9\,\delta_1 \tag{7-6g}$$

$$S(\mu,\delta_1,\delta_2) = \xi_3 - \xi_3(\mu+\psi) + \xi_4\,\delta_2 - [f_R + (f_L - f_R)(\mu+\psi) + \Delta(\delta_1+\delta_2)]\delta_2 \tag{7-6h}$$

$$U(q,\mu,\delta_1,\delta_2) = (f_L - f_R)q[1-2(\mu+\psi) + q) - \Delta(\delta_1+\delta_2)(\mu+\psi) + \Delta\delta_1 \tag{7-6i}$$

Theorem: Suppose $\gamma_1\,\gamma_2 \neq 0$ and $f_L \neq f_R$. Further suppose that $q(0) = 0$ (i.e., $p^L(0) = p_L(0)$, and define

$$t_o(\varepsilon) = \begin{cases} 0 & \text{if } \delta_1^2(0) + \delta_2^2(0) = 0 \\ -\dfrac{1}{\Gamma}\,\ell n\left[\dfrac{g^2(\varepsilon)}{\varepsilon}\right] & \text{otherwise .} \end{cases} \tag{7-7}$$

Suppose that $g(\varepsilon) = o(\varepsilon^{1/2})$ and $\varepsilon = O(g(\varepsilon))$, i.e.,

$$\lim_{\varepsilon\downarrow 0} \frac{g(\varepsilon)}{\varepsilon^{1/2}} = 0 \quad , \quad \lim_{\varepsilon\downarrow 0} \frac{\varepsilon}{g(\varepsilon)} \text{ exists} \quad . \tag{7-8}$$

Then for ε sufficiently small, there exists a positive constant C so that

$$\frac{\sup\limits_{t>t_o(\varepsilon)} E[q^2(t)]}{\sup\limits_{t>T} E[\mu^2(t)]} < C\,\varepsilon \tag{7-9}$$

for any $T>0$.

Let us make several comments about this result. First, the assumption that $\gamma_1 \gamma_2 \neq 0$ simply means that the 2-state aggregate process in Fig. 6-2 has a nontrivial ergodic distribution, while the condition $f_L \neq f_R$ states that the aggregated version of the measurements does contain at least some information. Also, from Eq. 7-7 we see that the need for $t_o(\varepsilon)$ is due entirely to boundary-layer effects caused by initial conditions on $\delta_1(t)$ and $\delta_2(t)$. Finally note that the assumption that $q(0) = 0$ is reasonable, as it certainly makes sense for the exact and approximate filters to be initialized identically.

What this theorem states is that for measurement quality in the range specified by Eq. 7-8, the mean square deviation is order ε in size when compared to the mean-square information content as measured by $\mu(t)$. The inclusion of the arbitrary time T in Eq. 7-9 in fact implies that this information content persists at a level far above that of the deviation $q(t)$. Note also that as should be clear from the proof, this result can be extended to the case of poorer measurement quality, i.e., $g(\varepsilon) = o(\varepsilon)$, with an appropriate change in the right-hand side of Eq. 7-9. The order $\varepsilon^{1/2}$, however, represents a critical cut-off point at the upper limit of measurement quality. This will also be seen in the results presented in the next section.

Proof of Theorem: For notational simplicity in what follows, we will use simplified notation of the type

$$R(t) = R(\mu(t), \delta_1(t), \delta_2(t)) \qquad (6\text{-}10)$$

and will refer to the specific arguments of such a function only when necessary. Also, let us note an extremely important fact: the processes for which

we are solving in Eq. 7-5 all represent differences between two probabilities, i.e., we know a priori that

$$|\mu(t)|,\ |\delta_1(t)|,\ |\delta_2(t)|,\ |q(t)| \leq 1 \quad . \tag{7-11}$$

Consequently, since the quantities in Eqs. 7-6b - 7-6i are all first- or second-order polynomials in these processes, we see that they are all bounded, i.e., there exists a positive number $K < \infty$ so that

$$|K(t)|,\ |M(t)|,\ |N(t)|,\ |P(t)|,\ |Q(t)|,\ |R(t)|,\ |S(t)|,\ |U(t)| \leq K \quad . \tag{7-12}$$

Our proof now proceeds by first bounding the sizes of $\delta_1(t)$ and $\delta_2(t)$, then $q(t)$, and finally $\mu(t)$. To begin, we note from Eq. 7-5b that

$$\delta_1(t) = e^{-\Lambda_1 t/\varepsilon}\delta_1(0) + \int_0^t e^{-\Lambda_1(t-\tau)/\varepsilon}P(\tau)d\tau + \frac{g(\varepsilon)}{\sqrt{\varepsilon}}\int_0^t e^{-\Lambda_1(t-\tau)}Q(\tau)dw(t) \quad . \tag{7-13}$$

Then, using the fact that

$$\Big(\sum_{i=1}^{M} x_i\Big)^2 \leq M\sum_{i=1}^{M} x_i^2 \tag{7-14}$$

we obtain

$$\begin{aligned} E[\delta_1^2(t)] \leq\ & 3\, e^{-2\Lambda_1 t/\varepsilon}E[\delta_1^2(0)] \\ & + 3\int_0^t\int_0^t e^{-\Lambda_1[(t-\tau)+(t-\sigma)]/\varepsilon}E[P(\tau)P(\sigma)]d\tau d\sigma \\ & + 3\,\frac{g^2(\varepsilon)}{\varepsilon}\int_0^t e^{-2\Lambda_1(t-\tau)/\varepsilon}E[Q^2(\tau)]d\tau \quad . \end{aligned} \tag{7-15}$$

Using Eqs. 7-11 and 7-12 we obtain

$$E[\delta_1^2(t)] \leq 3\, e^{-2\Lambda_1 t/\varepsilon} E[\delta_1^2(0)] + 3\, K^2 \left[\int_0^t e^{-\Lambda_1(t-\tau)/\varepsilon}\, d\tau \right]^2 + \frac{3\, g^2(\varepsilon) K^2}{\varepsilon} \int_0^t e^{-2\Lambda_1(t-\tau)/\varepsilon} d\tau \quad . \tag{7-16}$$

Then using the fact that

$$\int_0^t e^{-a(t-\tau)} d\tau = \int_0^t e^{-a\tau} d\tau \leq \int_0^\infty e^{-a\tau} d\tau = \frac{1}{a} \tag{7-17}$$

we obtain the bound

$$E[\delta_1^2(t)] \leq 3\, E[\delta_1^2(0)] e^{-2\Lambda_1 t/\varepsilon} + \frac{3\, K^2 \varepsilon^2}{\Lambda_1^2} + \frac{3\, K^2 g^2(\varepsilon)}{2\Lambda_1} \quad . \tag{7-18}$$

In an analogous fashion, we obtain the bound

$$E[\delta_2^2(t)] \leq 3\, E[\delta_2^2(0)] e^{-2\Lambda_2 t/\varepsilon} + \frac{3\, K^2 \varepsilon^2}{\Lambda_2^2} + \frac{3\, K^2 g^2(\varepsilon)}{2\Lambda_2} \quad . \tag{7-19}$$

Consider next q(t). From Eq. 7-5d and the fact that q(0) = 0, we have

$$q(t) = \int_0^t e^{-\Gamma(t-\tau)} K(\tau) d\tau + \frac{g^2(\varepsilon)}{\varepsilon} \int_0^t e^{-\Gamma(t-\tau)} M(\tau) d\tau + \frac{g(\varepsilon)}{\sqrt{\varepsilon}} \int_0^t e^{-\Gamma(t-\tau)} U(\tau) dw(\tau) \quad . \tag{7-20}$$

Using Eq. 6-14, once again we obtain

$$E[q^2(t)] \leq 3 \int_0^t \int_0^t e^{-\Gamma[(t-\tau)+(t-\sigma)]} E[K(\tau)K(\sigma)]d\tau d\sigma$$

$$+ 3 \frac{g^4(\varepsilon)}{\varepsilon^2} \int_0^t \int_0^t e^{-\Gamma[(t-\tau)+(t-\sigma)]} E[M(\tau)M(\sigma)]d\tau d\sigma \qquad (7\text{-}21)$$

$$+ 3 \frac{g^2(\varepsilon)}{\varepsilon} \int_0^t e^{-2\Gamma(t-\tau)} E[U^2(\tau)]d\tau \quad .$$

Then employing the inequality

$$|xy| \leq \frac{1}{2} (x^2 + y^2) \qquad (7\text{-}22)$$

we find

$$E[q^2(t)] \leq 3 \int_0^t \int_0^t e^{-\Gamma[(t-\tau)+(t-\sigma)]} E[K^2(\tau)]d\tau$$

$$+ 3 \frac{g^4(\varepsilon)}{\varepsilon^2} \int_0^t \int_0^t e^{-\Gamma[(t-\tau)+(t-\sigma)]} E[M^2(\tau)]d\tau \qquad (7\text{-}23)$$

$$+ 3 \frac{g^2(\varepsilon)}{\varepsilon} \int_0^t e^{-2\Gamma(t-\tau)} E[U^2(\tau)]d\tau \quad .$$

Again using Eq. 7-17 we obtain

$$E[q^2(t)] \leq \frac{3}{\Gamma} \int_0^t e^{-\Gamma(t-\tau)} E[K^2(t)]d\tau$$

$$+ \frac{3\, g^4(\varepsilon)}{\Gamma\varepsilon^2} \int_0^t e^{-\Gamma(t-\tau)} E[M^2(\tau)]d\tau \qquad (7\text{-}24)$$

$$+ \frac{3\, g^2(\varepsilon)}{\varepsilon} \int_0^t e^{-2\Gamma(t-\tau)} E[U^2(\tau)]d\tau \quad .$$

Let us now examine the first integral on the right-hand side of Eq. 7-24. From Eqs. 7-6b, 7-14, 7-18, and 7-19 we see that

$$\begin{aligned} E[K^2(t)] &\leq 2\,\eta_1^2\,E[\delta_1^2(t)] + 2\,\eta_2^2\,E[\delta_2^2(t)] \\ &\leq A_1\,E[\delta_2^2(0) + \delta_2^2(0)]e^{-2\,\Lambda t/\varepsilon} \\ &\quad + A_2\,\varepsilon\,e^{-\Lambda t/\varepsilon} + A_3\,\varepsilon^2 + A_4\,g^2(\varepsilon) \end{aligned} \tag{7-25}$$

where $\Lambda = \min(\Lambda_1,\Lambda_2)$ and A_1, A_2, A_3, A_4 are appropriate postive constants. Also, from Eq. 7-8 for ε sufficiently small*

$$\begin{aligned} E[K^2(t)] &\leq A_1\,E[\delta_1^2(0) + \delta_2^2(0)]e^{-2\Lambda t/\varepsilon} \\ &\quad + A_2\,\varepsilon\,e^{-\Lambda t/\varepsilon} + A_5\,g^2(\varepsilon) \end{aligned} \tag{7-26}$$

for same constant A_5.

Using Eq. 7-26 we then have that

$$\begin{aligned} &\int_0^t e^{-\Gamma(t-\tau)}E[K^2(\tau)]d\tau \\ &\quad \leq \left[A_1\{E[\delta_1^2(0)] + E[\delta_2^2(0)] + A_2\varepsilon\}\,\frac{\varepsilon\,e^{-\Gamma t}}{2\Lambda-\varepsilon\Gamma}\left(1 - e^{(\Gamma-\Lambda/\varepsilon)t}\right)\right. \\ &\qquad \left. + \frac{A_5\,g^2(\varepsilon)}{\Gamma}\left(1 - e^{-\Gamma t}\right)\right. \\ &\quad \leq \frac{A_1\{E[\delta_1^2(0)] + E[\delta_2^2(0)]\}\varepsilon\,e^{-\Gamma t}}{2\Lambda} + A_6\,g^2(\varepsilon) \end{aligned} \tag{7-27}$$

where we have again used Eq. 7-8 and are assuming ε sufficiently small.

*This is a typical place where we use the fact that $g(\varepsilon)$ is no smaller than order ε. If $g(\varepsilon) = O(\varepsilon)$, the last term in Eq. 7-26 would be $A_5\,\varepsilon^2$.

We next examine the second integral on the right-hand side of Eq. 7-24. In this case we will be a bit more careful. Specifically, from Eqs. 7-6c and 7-14 we have that

$$\int_0^t e^{-\Gamma(t-\tau)} E[M^2(\tau)]d\tau \leq 3(f_L - f_R)^2 \int_0^t e^{-\Gamma(t-\tau)} E[q^2(\tau)]d\tau + 3\Delta^2 \int_0^t e^{-\Gamma(t-\tau)} \{E[\delta_1^2(\tau) + \delta_2^2(\tau)\}d\tau \quad . \tag{7-28}$$

In a manner analogous to the calculations in Eqs. 7-25 - 7-27, we can bound the second integral on the right-hand side of Eq. 7-30 in exactly the same form as in Eq. 7-27. Thus for ε sufficiently small

$$\int_0^t e^{-\Gamma(t-\tau)} E[M^2(\tau)]d\tau \leq 3(f_L - f_R)^2 \int_0^t e^{-\Gamma(t-\tau)} E[q^2(\tau)]d\tau + A_7\{E[\delta_1^2(0)] + E[\delta_2^2(0)]\}\varepsilon\, e^{-\Gamma t} + A_8\, g^2(\varepsilon) \quad . \tag{7-29}$$

Examining next the third integral on the right-hand side of Eq. 7-24 and using Eqs. 7-6i and 7-14, we have

$$\int_0^t e^{-2\Gamma(t-\tau)} E[U^2(t)]d\tau \leq 3 \int_0^t e^{-2\Gamma(t-\tau)} E[d_1^2(\tau)q^2(\tau)]d\tau + 3 \int_0^t e^{-2\Gamma(t-\tau)} E[d_2^2(\tau)\ \delta_1^2(\tau) + d_3^2(\tau)\delta_2^2(\tau)]d\tau \tag{7-30}$$

where

$$d_1(t) = (f_L - f_R)[1-2(\mu(t) + \psi) + q(t)] \tag{7-31a}$$

$$d_2(t) = \Delta(1 - \mu(t) - \Psi) \tag{7-31b}$$

$$d_3(t) = \Delta(\mu(t) + \psi) \quad . \tag{7-31c}$$

Thanks to Eqs. 7-11 and 7-12, $d_1(t)$, $d_2(t)$, and $d_3(t)$ are bounded so that we can obtain a simple bound

$$\begin{aligned}\int_0^t e^{-2\Gamma(t-\tau)} E[U^2(\tau)]d\tau \leq A_9 \int_0^t e^{-2\Gamma(t-\tau)} E[q^2(\tau)]d\tau \\ + A_{10} \int_0^t e^{-2\Gamma(t-\tau)} E\{\delta_1^2(\tau) + \delta_2^2(\tau)\}d\tau\end{aligned} \tag{7-32}$$

and then again in a manner analogous to Eqs. 7-25 - 7-27, we can obtain the following bound for ε sufficiently small

$$\begin{aligned}\int_0^t e^{-2\Gamma(t-\tau)} E[U^2(\tau)]d\tau \leq A_9 \int_0^t e^{-2\Gamma(t-\tau)} E[q^2(\tau)]d\tau \\ + A_{11}\{E[\delta_1^2(0)] + E[\delta_2^2(0)]\}\varepsilon\, e^{-2\Gamma t} \\ + A_{12}\, g^2(\varepsilon) \quad .\end{aligned} \tag{7-33}$$

Combining Eqs. 7-24, 7-27, 7-29 and 7-33 (and performing a few additional, straightforward bounds), we obtain a bound of the following form for ε sufficiently small

$$\begin{aligned}E[q^2(t)] \leq A_{13}\, g^2(\varepsilon) + A_{14}\, E\{\delta_1^2(0) + \delta_2^2(0)\}\varepsilon\, e^{-\Gamma t} \\ + A_{15}\, \frac{g^4(\varepsilon)}{\varepsilon^2} \int_0^t e^{-\Gamma(t-\tau)} E[q^2(\tau)]d\tau \\ + A_{16}\, \frac{g^2(\varepsilon)}{\varepsilon} \int_0^t e^{-2\Gamma(t-\tau)} E[q^2(\tau)]d\tau \quad .\end{aligned} \tag{7-34}$$

We can now use Eq. 7-34 to obtain a first bound. Specifically, let

$$M = \sup_{t \geq 0} E[q^2(t)] \quad . \tag{7-35}$$

Then using Eqs. 7-17, 7-34, and 7-35, we obtain

$$M \leq A_{13}\, g^2(\varepsilon) + A_{14}\, E\{\delta_1^2(0) + \delta_2^2(0)\}\, \varepsilon + \frac{A_{15}\, M}{\Gamma}\, \frac{g^4(\varepsilon)}{\varepsilon^2} + \frac{A_{16}\, M}{2\Gamma}\, \frac{g^2(\varepsilon)}{\varepsilon} \tag{7-36}$$

i.e.,

$$M \leq \frac{A_{13}\, g^2(\varepsilon) + A_{14}\, E\{\delta_1(0) + \delta_2^2(0)\}\varepsilon}{1 - \frac{A_{15}\, M}{\Gamma}\, \frac{g^4(\varepsilon)}{\varepsilon^2} - \frac{A_{16}\, M}{2\Gamma}\, \frac{g^2(\varepsilon)}{\varepsilon}} \leq A_{17}\, \varepsilon \tag{7-37}$$

for ε sufficiently small. Using this bound in the two integrands in Eq. 7-34 and using Eqs. 7-17 and 7-37, we see that

$$\begin{aligned} E[q^2(t)] &\leq A_{13}\, g^2(\varepsilon) + A_{14}\, E\{\delta_1^2(0) + \delta_2^2(0)\}\varepsilon\, e^{-\Gamma t} \\ &\quad + \frac{A_{15}\, A_{17}}{\Gamma}\, \frac{g^4(\varepsilon)}{\varepsilon} + \frac{A_{16}\, A_{17}}{2\,\Gamma}\, g^2(\varepsilon) \\ &\leq A_{18}\, g^2(\varepsilon) + A_{14}\, E\{\delta_1^2(0) + \delta_2^2(0)\}\varepsilon\, e^{-\Gamma t} \end{aligned} \tag{7-38}$$

where the last inequality follows from Eq. 7-8 and is valid for ε sufficiently small. From this we immediately see that

$$\sup_{t \geq t_o(\varepsilon)} E[q^2(t)] \leq A_{14}\, g^2(\varepsilon) \tag{7-39}$$

where $t_o(\varepsilon)$ is given by Eq. 7-7.

Now we turn our attention to $\mu(t)$. Without loss of generality, assume $\mu(0) = 0$ (this is the worst case coresponding to no prior information). Then from Eq. 7-5a

$$\mu(t) = \int_0^t e^{-\Gamma(t-\tau)} K(\tau) d\tau + \frac{g(\varepsilon)}{\sqrt{\varepsilon}} \int_0^t e^{-\Gamma(t-\tau)} N(\tau) dw(\tau) \quad . \tag{7-40}$$

First we look at the mean of $\mu(t)$:

$$E[\mu(t)] = \int_0^t e^{-\Gamma(t-\tau)} E[K(\tau)] d\tau \quad . \tag{7-41}$$

Then

$$|E[\mu(t)]| \leq \int_0^t e^{-\Gamma(t-\tau)} |E[K(\tau)]| d\tau \leq \int_0^t e^{-\Gamma(t-\tau)} \{E[K^2(\tau)]\}^{1/2} d\tau \quad . \tag{7-42}$$

Then, using Eq. 7-26 and the fact that for $x_i \geq 0$

$$\left(\sum_{i=1}^{M} x_i \right)^{1/2} \leq \sum_{i=1}^{M} (x_i)^{1/2} \tag{7-43}$$

we find that for ε sufficiently small

$$\begin{aligned} |E[\mu(t)]| &\leq \int_0^t e^{-\Gamma(t-\tau)} [B_1 \, e^{-\Lambda\tau/\varepsilon} + B_2 \, g(\varepsilon)] d\tau \\ &= \frac{B_1 \, \varepsilon \, e^{-\Gamma t}}{\Lambda - \varepsilon \, \Gamma} \left(1 - e^{(\Gamma - \Lambda/\varepsilon)t}\right) + \frac{B_2 \, g(\varepsilon)}{\Gamma} \left(1 - e^{-\Gamma t}\right) \\ &\leq B_3 \, g(\varepsilon) \end{aligned} \tag{7-44}$$

where the B_i are positive constants and again we have used Eq. 7-8 in the last step.

Next we examine $E[\mu^2(t)]$. To do this, let

$$x = \int_0^t e^{-\Gamma(t-\tau)} K(\tau) d\tau \tag{7-45a}$$

$$y = \frac{g(\varepsilon)}{\sqrt{\varepsilon}} \int_0^t e^{-\Gamma(t-\tau)} N(\tau) dw(\tau) \quad . \tag{7-45b}$$

Note also that from the analysis in Eqs. 7-22 - 7-27 together with Eq. 7-43, we can deduce that

$$\begin{aligned} \{E(x^2)\}^{1/2} &< \left\{ \frac{A_1 \, E[\delta_1^2(0) + \delta_2^2(0)] \varepsilon \, e^{-\Gamma t}}{2 \Lambda} + A_6 \, g^2(\varepsilon) \right\}^{1/2} \\ &< B_4 [\delta_1^2(0) + \delta_2^2(0)]^{1/2} \sqrt{\varepsilon} \, e^{-\Gamma t/2} + B_5 \, g(\varepsilon) \quad . \end{aligned} \tag{7-46}$$

Next, let us obtain an upper bound on $E(y^2)$. Specifically, using Eqs. 7-12 and 7-17

$$\begin{aligned} E(y^2) &= \frac{g^2(\varepsilon)}{\varepsilon} \int_0^t e^{-2\Gamma(t-\tau)} E[N(\tau)^2] d\tau \\ &< \frac{K^2 \, g^2(\varepsilon)}{2 \, \Gamma \, \varepsilon} \triangleq B_6 \, \frac{g^2(\varepsilon)}{\varepsilon} \quad . \end{aligned} \tag{7-47}$$

Now, since $\mu(t) = x+y$, we can obtain the lower bound

$$\begin{aligned} E[\mu^2(t)] &= E[(x+y)^2] \\ &> E(y^2) - 2\{E(x^2)\}^{1/2} \{E(y^2)\}^{1/2} \\ &> E(y^2) - 2\sqrt{B_6} \, \frac{g(\varepsilon)}{\sqrt{\varepsilon}} \{B_4 \, E[\delta_1^2(0) + \delta_2^2(0)] \sqrt{\varepsilon} \, e^{-\Gamma t/2} + B_5 \, g(\varepsilon)\} \end{aligned} \tag{7-48}$$

We now proceed to find a lower bound on $E(y^2)$. To begin, note that for $t>1$

$$E(y^2) > \frac{g^2(\varepsilon)}{\varepsilon} \int_0^t e^{-2\Gamma(t-\tau)} E[N^2(\tau)]d\tau \quad . \tag{7-49}$$

Now from Eq. 7-6d

$$N(t) = C_0(t) + C_1(t)\ \mu(t) + C_2\ \mu^2(t) \tag{7-50}$$

where

$$C_0(t) = [(f_L - f_R)\ (1-\psi)\ - \Delta(\delta_1(t) + \delta_2(t))]\ \psi + \Delta\delta_1(t) \tag{7-51a}$$

$$C_1(t) = (f_L - f_R)\ (1-2\psi) - \Delta[\delta_1(t) + \delta_2(t)] \tag{7-51b}$$

$$C_2 = f_R - f_L \quad . \tag{7-51c}$$

Substituting Eq. 7-50 into Eq. 7-49 and discarding several nonnegative terms, we obtain

$$\begin{aligned} E[y^2] > \frac{g^2(\varepsilon)}{\varepsilon} \int_0^t e^{-2\Gamma(t-\tau)} E\{&C_0^2(\tau) + 2\ C_0(\tau)\ C_1(\tau)\ \mu(\tau) \\ &+ 2\ C_0(\tau)\ C_2\ \mu^2(\tau) + 2\ C_1(\tau)\ C_2\ \mu^3(\tau)\}d\tau \quad . \end{aligned} \tag{7-52}$$

From Eq. 7-51a

$$\begin{aligned} C_0^2(t) &= (f_L - f_R)^2\ (1-\psi)^2\ \psi^2 + 2(f_L - f_R)\ (1-\psi)\ \psi\Delta[\delta_1(t) - \psi\ \delta_1(t) - \psi\ \delta_2(t)] \\ &\quad + \Delta^2[\delta_1(t) - \psi\ \delta_1(t) + \psi\ \delta_2(t)]^2 \\ &> B_7 - B_8|\delta_1(t) - \psi\ \delta_1(t) - \psi\ \delta_2(t)| \end{aligned} \tag{7-53}$$

where $B_7\ (f_L - f_R)^2(1-\psi)^2\ \psi^2$ and $B_8 = 2|(f_L - f_R)\ (1-\psi)\psi\Delta|$ are positive constants. Thus

$$
\begin{aligned}
E[C_0^2(t)] &\geq B_7 - B_8\, E[|\delta_1(t) - \psi\, \delta_1(t) - \psi\, \delta_2(t)|] \\
&\geq B_7 - B_8\{E[\delta_1(t) - \psi\, \delta_1(t) - \psi\, \delta_2(t)]^2\}^{1/2} \\
&\geq B_7 - B_8\{B_9\, E[\delta_1^2(0) + \delta_2^2(0)]^{1/2}\, e^{-\Lambda t/\varepsilon} \\
&\quad + B_{10}\, \varepsilon^{1/2}\, e^{-\Lambda t/2\varepsilon} + B_{11}\, g(\varepsilon)\}
\end{aligned}
\tag{7-54}
$$

where the last inequality is obtained in the same manner as Eq. 7-26 followed by an application of Eq. 7-43. Note that is we define

$$
t_1(\varepsilon) = \begin{cases} 0 & \text{if } \delta_1^2(0) + \delta_2^2(0) = 0 \\ \dfrac{-\varepsilon\, \ell n\, g(\varepsilon)}{\Lambda} & \text{otherwise} \end{cases} .
\tag{7-55}
$$

Then

$$
E[C_0^2(t)] \geq B_{12} > 0 \text{ for } t \geq t_1(\varepsilon) .
\tag{7-56}
$$

Note that for ε small enough $t_1(\varepsilon) = t_0(\varepsilon) = 0$ or $t_1(\varepsilon) << t_0(\varepsilon)$. Now, from Eqs. 7-52 and 7-56 we have that for $t \geq \max(t_1(\varepsilon),1)$

$$
\begin{aligned}
E[y^2] \geq B_{12}\, \frac{g^2(\varepsilon)}{\varepsilon} - 2\, \frac{g^2(\varepsilon)}{\varepsilon} \int_{t-1}^{t} E\{&C_0(\tau)C_1(\tau)\mu(\tau) + C_0(\tau)C_2\mu^2(\tau) \\
&+ C_1(\tau)C_2\mu^3(\tau)\}d\tau .
\end{aligned}
\tag{7-57}
$$

Since $C_0(t)$ and $C_1(t)$ are bounded and $|\mu(t)| < 1$, we deduce that

$$
E(y^2) \geq B_{12}\, \frac{g^2(\varepsilon)}{\varepsilon} - B_{13}\, \frac{g^2(\varepsilon)}{\varepsilon} \int_{t-1}^{t} \{|E[\mu(\tau)]| + E[\mu^2(\tau)]\}d\tau .
\tag{7-58}
$$

Then from Eq. 7-44 we have that for $t \geq \max(t_1(\varepsilon),1)$

$$E(y^2) \geq B_{12} \frac{g^2(\varepsilon)}{\varepsilon} - B_3 B_{13} \frac{g^3(\varepsilon)}{\varepsilon} - B_{13} \frac{g^2(\varepsilon)}{\varepsilon} \int_{t-1}^{t} E[\mu^2(\tau)]d\tau$$

$$\geq B_{14} \frac{g^2(\varepsilon)}{\varepsilon} - B_{13} \frac{g^2(\varepsilon)}{\varepsilon} \int_{t-1}^{t} E[\mu^2(\tau)]d\tau \qquad (7\text{-}59)$$

where $B_{14} > 0$ and ε is sufficiently small.

Combining Eqs. 7-59 and 7-48, we have that for $t \geq \max(t_1(\varepsilon),1)$

$$E[\mu^2(t)] \geq B_{14} \frac{g^2(\varepsilon)}{\varepsilon} - B_{13} \frac{g^2(\varepsilon)}{\varepsilon} \int_{t-1}^{t} E[\mu^2(\tau)]d\tau$$

$$- B_{15} E[\delta_1^2(0) + \delta_2^2(0)] g(\varepsilon) e^{-\Gamma t/2} - B_{16} \frac{g^2(\varepsilon)}{\sqrt{\varepsilon}} . \qquad (7\text{-}60)$$

Then for $t \geq \max(t_0(\varepsilon),1)$ and ε sufficiently small

$$E[\mu^2(t)] \geq B_{17} \frac{g^2(\varepsilon)}{\varepsilon} - B_{13} \frac{g^2(\varepsilon)}{\varepsilon} \int_{t-1}^{t} E[\mu^2(\tau)]d\tau \qquad (7\text{-}61)$$

where $B_{17} > 0$.

Now, let T be any time ≥ 0 and define

$$N = \sup_{t \geq T} E[\mu^2(t)] . \qquad (7\text{-}62)$$

Then, using Eq. 7-61 we obtain

$$N \geq \sup_{t \geq \max(t_0(\varepsilon),T+1)} E[\mu^2(t)] \geq B_{17} \frac{g^2(\varepsilon)}{\varepsilon} - B_{13} N \frac{g^2(\varepsilon)}{\varepsilon} \qquad (7\text{-}63)$$

from which we deduce that

$$N \geq B_{18} \frac{g^2(\varepsilon)}{\varepsilon} \qquad (7\text{-}64)$$

with $B_{18} > 0$ and for ε sufficiently small. Combining Eqs. 7-39 and 7-64 completes the proof of the theorem.

In the next section we present simulation results that corroborate this result and that in fact suggest several other, stronger results. Several conjectures are presented in the conclusions.

SECTION 8

MONTE CARLO ANALYSIS OF FILTERING PROBLEMS

In this section, the simulation of the filtering techniques proposed in the previous sections are presented. Both the techniques used to simulate the processes of interest and the filters themselves are described, followed by the results. The goals of the simulations were two-fold. First they were intended to provide support for results and conjectures regarding the quality of various filtering approximation schemes. Secondly, they were intended to clarify the types of phenomena that can occur in filtering problems involving processes with various time-scale separations, noise levels and information rates. As we will see, the results obtained from the simuations support the results of the previous section as well as suggesting several additional conjectures.

All simulations were carried out on a PC's Unlimited 286 personal computer. Simulation routines were written using Turbo Pascal.

8.1 Simulation Techniques

Before discussing specific simulations we describe the techniques required to simulate both the processes of interest and the filters themselves. The techniques of interest relate to the simulation of Wiener Processes, Markov Processes and stochastic differential equations.

8.1.1 Noise Processes

In the simulations we need to generate increments of Brownian motion processes. These increments were generated by manipulating samples obtained from a uniform random number generator. The sample values were drawn from a uniform distribution on [0,1], summed and scaled to approximate a value drawn from a Gaussian distribution. The scaling factor Q was given by

$$Q = \left(\frac{12\ I\ \Delta t}{N}\right)^{i/2} \tag{8-1}$$

where I is the intensity of the noise process, Δt is the time increment between sample times in the simulation, and N is the number of values drawn from a uniform distribution in order to generate a simple, approximately Gaussian random variable. A nominal value of N = 10 was chosen for these simulations.

8.1.2 Markov Processes

For each of the filtering simulations described in Section 6, a sample path for the 4-state Markov Process of Fig. 6-1 is required. The generation of a sample path for a general continuous-time Markov process can be accomplished by generating the sequence of successive states that are visited and the time intervals between the transitions from state-to-state in the sequence. Specifically, let x_N be the state after the N-th jump and t_N the time of the N-th jump, we can define τ_N to be the N-th holding time for the system and p_{ij} the conditional jump probabilities where:

$$\tau_N = t_N - t_{N-1} \tag{8-2}$$

$$p_{ij} = \Pr\{x_N = j \mid x_{N-1} = i\} \tag{8-3}$$

The generation of the sample path then proceeds as follows. Suppose that we have generated $x_{N-1} = i$ at time t_{N-1}. Then we generate two conditionally independent random variables (conditional on $x_{N-1} = i$), namely x_N and τ_N. The required density for τ_N is given by

$$p(\tau_N \mid x_N = i) = \lambda_{T_i} \exp\{-\lambda_{T_i} \tau_N\} \tag{8-4}$$

where

$$\lambda_{ij} = \text{the transition rate from state i to state j,}$$

and

$$\lambda_{T_i} = \sum_{j \neq i} \lambda_{ij} \quad . \tag{8-5}$$

The technique used to generate the required random variables was to generate a value z' from a uniform distribution on [0,1]. A new value z is then calculated as

$$z = - \lambda_{T_i}^{-1} \ell n \ (z') \tag{8-6}$$

so that

$$p_z(z) = \lambda_{T_i} \exp\{- \lambda_{T_i} z\} \quad . \tag{8-7}$$

To decide on the order of states entered for the sample path, another variable Y_N' was drawn from a uniform distribution on [0,1]. A series of thresholds were then calculated as

$$\tilde{Y}_o = 0 \tag{8-8a}$$

$$\tilde{Y}_k = \sum_{\substack{k=0 \\ k \neq i}}^{N-1} \frac{\lambda_{ik}}{\lambda_{T_i}} \quad . \tag{8-8b}$$

The state after the N-th jump is then obtained as

$$x_N = j \quad \text{if} \quad \tilde{Y}_{j-1} < Y_N' < \tilde{Y}_j \quad , \tag{8-9}$$

which corresponds to

$$p_{ij} = \begin{cases} \dfrac{\lambda_{ik}}{\lambda_{T_i}} & ; \quad k \neq i \\ 0 & k = i \end{cases} \quad . \tag{8-10}$$

8.2 SIMULATION OF DIFFERENTIAL EQUATIONS

In order to simulate the performance of the various filtering approaches, it is necessary to integrate the differential equations describing the filters. We consider two cases:

1. Filtering with noise present (Stochastic Differential Equations)
2. Filtering in the noiseless environment

8.2.1 Simulation in a Noisy Environment

In the cases where we observe a process with additive noise, we generally obtain filtering equations of the form

$$d\underline{p} = \underline{F}(\underline{p})dt + \underline{G}(\underline{p})d\nu \tag{8-11}$$

where $\underline{p}$ is a vector of probabilities, ν is an innovations process with unit intensity, and $\underline{F}$ and $\underline{G}$ are column vectors which we denote by

$$\underline{F} = \begin{bmatrix} f_1(\underline{p}) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ f_n(\underline{p}) \end{bmatrix} \qquad \underline{G} = \begin{bmatrix} g_1(\underline{p}) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ g_n(\underline{p}) \end{bmatrix} \quad . \tag{8-12}$$

In order to simulate the filter (and the process of interest) on a digital computer, it is necessary to discretize time into segments of length Δt and propagate the equations over each interval. The method that was chosen to propagate the differential equation over each interval is a Runge-Kutta scheme adapted to the problem of nonlinear stochastic differential equations [8].

This technique integrates a differential equation from time t_N to time $t_{N+1} = t_N + \Delta t$ as follows:

$$\underline{p}'(t_{N+1}) = \underline{p}(t_N) + \tilde{\underline{F}}(p(t_N))\Delta t + \underline{G}(\underline{p}(t_N))\ \Delta\nu \tag{8-13a}$$

$$\begin{aligned}\underline{p}(t_{N+1}) &= \underline{p}(t_N) + \left(\frac{\tilde{\underline{F}}(\underline{p}(t_N)) + \tilde{\underline{F}}(\underline{p}'(t_{N+1}))}{2}\right)\Delta t \\ &\quad + \left(\frac{\underline{G}(\underline{p}(t_N)) + \underline{G}(\underline{p}'(t_{N+1}))}{2}\right)\Delta\nu\end{aligned} \tag{8-13b}$$

where $\tilde{F}$ is a modified function required so that our simulation approximates Eq. 8-11. Specifically, it is known that as $\Delta t \to 0$, the solution of Eq. 8-13 converges to the solution of the differential equation

$$d\underline{p}(t) = \left[\tilde{\underline{F}} + \frac{1}{2}\frac{\partial \underline{G}(\underline{p})}{\partial \underline{p}}\underline{G}(\underline{p})\right]dt + \underline{G}(\underline{p})d\nu = \underline{F}\ dt + \underline{G}\ d\nu\ . \tag{8-14}$$

We therefore require that

$$\tilde{\underline{F}}(p) = \underline{F}(p) - \frac{1}{2}\frac{\partial \underline{G}(\underline{p})}{\partial \underline{p}}\underline{G}\ (p) \tag{8-15}$$

where

$$\frac{\partial \underline{G}(\underline{p})}{\partial \underline{p}} = \begin{bmatrix} \frac{\partial G_1(\underline{p})}{\partial p_1} & \cdots & \frac{\partial G_1(\underline{p})}{\partial p_N} \\ \vdots & & \vdots \\ \frac{\partial G_N(\underline{p})}{\partial p_1} & \cdots & \frac{\partial G_N(\underline{p})}{\partial p_N} \end{bmatrix} \tag{8-16}$$

Therefore, the matrix of expressions $\tilde{F}$ that we use in Eq. 8-13, differs from the matrix $\underline{F}$ in the equation describing the process we wish to simulate. We proceed by providing the equations that were used to simulate the various filters conjectured for the noise-corrupted observation model.

I. EXACT OPTIMAL NONLINEAR FILTER

The original equations were defined in Eqs. 6-8 and 6-9 as

$$dp^L(t) = \varepsilon[\gamma_2 - (\gamma_1+\gamma_2)p^L(t)]dt + \varepsilon[-\eta_1\ \delta_1(t) + \eta_2\ \delta_2(t)]dt + g(\varepsilon)\ [f_L - \hat{h}(t)]p^L(t)d\nu(t) + g(\varepsilon)\ \Delta\delta_1(t)\ d\nu(t)$$

$$d\delta_1(t) = -\Lambda_1\delta_1(t)dt + \varepsilon[\eta_5 - \eta_3\delta_1(t) + (\eta_4 - \eta_5)p^L(t) + \eta_6\ \delta_2(t)]dt + g(\varepsilon)[\xi_1\ p^L(t) + \xi_2\ \delta_1(t) - \hat{h}(t)\ \delta_1(t)]d\nu(t)$$

$$d\delta_2(t) = -\Lambda_2\ \delta_2(t)dt + \varepsilon[\eta_8 - \eta_4\ \delta_2(t) + (\eta_{10} - \eta_8)p^L(t) + \eta_9\ \delta_1(t)]dt + g(\varepsilon)[\xi_3 - \xi_3\ p^L(t) + \xi_4\ \delta_2(t) - \hat{h}(t)\ \delta_2(t)]d\nu(t)\ .$$

Casting this in the format of Eq. 8-14 we obtain

$$\underline{F}(\underline{p}) = \begin{bmatrix} \varepsilon[\gamma_2 - (\gamma_1+\gamma_2)p^L(t) + \eta_2\,\delta_2(t) - \eta_1\,\delta_1(t)] \\ -(\Lambda_1 + \varepsilon\,\eta_3)\delta_1(t) + \varepsilon[(\eta_4 - \eta_5)p^L(t) + \eta_5 + \eta_6\,\delta_2(t) \\ -(\Lambda_2 + \varepsilon\,\eta_7)\delta_2(t) + \varepsilon[(\eta_{10} - \eta_8)p^L(t) + \eta_8 + \eta_9\,\delta_1(t)] \end{bmatrix} \tag{8-17a}$$

$$\underline{G}(\underline{p}) = g(\varepsilon) \begin{bmatrix} f_L - \hat{h}(t))p^L(t) + \Delta\delta_1(t) \\ \xi_1\,p^L(t) + (\xi_2 - \hat{h})\delta_1 \\ \xi_3(1 - p^L(t)) + (\xi_4 - \hat{h})\delta_2 \end{bmatrix} \tag{8-17b}$$

$$\frac{\partial \underline{G}(p)}{\partial \underline{p}} = g(\varepsilon) \begin{bmatrix} (f_L - \hat{h}) - (f_L - f_R)p^L(t) & \Delta(1 - p^L(t)) & -\Delta p^L(t) \\ \xi_1 - (f_L - f_R)\delta_1(t) & \xi_2 - (\hat{h} + \Delta\delta_1(t)) & -\Delta\delta_1(t) \\ -\xi_3 - (f_L - f_R)\delta_2(t) & -\Delta\delta_2(t) & \xi_4 - (\hat{h} + \Delta\delta_2(t)) \end{bmatrix} \tag{8-18}$$

and letting $\tilde{\underline{F}} = \underline{F}(p) - \frac{1}{2}\frac{\partial \underline{G}(p)}{\partial \underline{p}} \cdot \underline{G}(\underline{p})$, we can apply the propagation equation given by Eq. 8-14. The increment of time, Δt, was restricted to be, at most, 10 percent of the mean time between the fastest transitions of the Markov Chain.

II. TWO-STATE APPROXIMATE FILTER

This filter was described in subsection 6.1 and satisfies Eqs. 6-16 and 6-14. It was simulated in exactly the same manner as the exact filter described above except that the calculations are simplified because $\delta_1(t)$, $\delta_2(t)$ are assumed to be zero. We therefore obtain

$$F(p_L(t)) = \varepsilon(\gamma_2 - (\gamma_1+\gamma_2)p_L(t)) \tag{8-19a}$$

$$G(p_L(t)) = g(\varepsilon)\,(f_L-f_R)\,p_L(t)\,(1 - p_L(t)) \tag{8-19b}$$

$$\frac{\partial}{\partial p_L} G(p_L(t)) = g(\varepsilon)\,(f_L-f_R)\,(1 - 2\,p_L(t)) \tag{8-19c}$$

and

$$\tilde{F}(p_L(t)) = \varepsilon(\gamma_2 - (\gamma_1+\gamma_2)\,p_L(t)) - \frac{1}{2}\,g^2(\varepsilon)\,(f_L-f_R)^2\,p_L(t)\,(1 - p_L(t))\,(1 - 2\,p_L(t)) \,. \tag{8-19d}$$

Using these equations, Eq. 8-14 was applied once again.

III. FRONT-END/BACK-END PROCESSOR

Simulations were performed for the front-end/back-end processor for which the front end calculated the likelihood ratio, L(t) using the equations given by Eq. 6-26, with $d\nu_R$, $d\nu_L$ given by Eq. 6-25b. Once again, Eq. 6-26 is a stochastic differential equation driven by white noise and requires a corrected drift term for the Runge Kutta technique. If we define $\underline{p}(t)$ as

$$p(t) = \begin{bmatrix} L(t) \\ Q_{LT}(t) \\ Q_{RT}(t) \end{bmatrix} , \tag{8-20}$$

where L(t), Q_{LT}, Q_{RT} are defined as in Eq. 6-26, we can use our previous formulation to simulate

$$d\underline{p}(t) = \underline{F}(\underline{p}(t))dt + \underline{G}(\underline{p}(t)) \begin{bmatrix} d\nu_R(t) \\ d\nu_L(t) \end{bmatrix}$$

where

$$\underline{F}(\underline{p}) = \begin{bmatrix} 0 \\ \lambda_2 - (\lambda_1+\lambda_2)\ Q_{LT}(t) \\ \lambda_4 - (\lambda_3+\lambda_4)\ Q_{RT}(t) \end{bmatrix} \tag{8-21a}$$

and

$$\underline{G}(\underline{p}(t)) = \left[G_R(\underline{p}) \vdots G_L(\underline{p}) \right] = \left[\begin{array}{c:c} g(\epsilon)(\hat{h}_L(t)-\hat{h}_R(t))\ L(t) & 0 \\ 0 & g(\epsilon)\ (\alpha - h_L(t))\ Q_{LT} \\ g(\epsilon)\ (\alpha - \hat{h}_R(t))\ Q_{RT} & 0 \end{array} \right], \tag{8-21b}$$

with $\underline{G}$ segmented into two column vectors, one associated with each of the two processes ν_L, ν_R. Note that ν_L and ν_R are not innovations processes. Rather we have that

$$\begin{bmatrix} d\nu_R(t) \\ d\nu_L(t) \end{bmatrix} = \begin{bmatrix} dy(t) - g(\epsilon)\ \hat{h}_L(t)dt \\ dy(t) - g(\epsilon)\ \hat{h}_R(t)dt \end{bmatrix} \tag{8-22}$$

where

$$\hat{h}_L(t) = \beta + (\alpha-\beta)\ Q_{LT}(t) \tag{8-23a}$$

$$\hat{h}_R(t) = \beta + (\alpha-\beta)\ Q_{RT}(t) \tag{8-23b}$$

and the true innovations is given by

$$d\nu(t) = dy(t) - g(\varepsilon)\,\hat{h}(t)dt \quad .$$

Since $d\nu_R$ and $d\nu_L$ differs from $d\nu$ only by a drift term, we can simply obtain the correction terms corresponding to ν_R and ν_L and add them. Specifically

$$\frac{\partial G_R}{\partial \underline{p}} = \begin{bmatrix} g(\varepsilon)\,(\hat{h}_L(t) - \hat{h}_R(t)) & g(\varepsilon)(\alpha-\beta)L(t) & -g(\varepsilon)(\alpha-\beta)L(t) \\ 0 & 0 & 0 \\ 0 & 0 & g(\varepsilon)(\alpha-\beta)(1-2Q_{RT}(t)) \end{bmatrix} \tag{8-24a}$$

$$\frac{\partial G_L}{\partial \underline{p}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & g(\varepsilon)(\alpha-\beta)(1-2Q_{LT}(t)) & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{8-24b}$$

Finally, we obtain $\tilde{F}(\underline{p})$ for simulation purposes by summing the correction terms associated with $d\nu_L$ and $d\nu_R$ to obtain:

$$\tilde{F}(\underline{p}) = \begin{bmatrix} -Y_2\, g^2(\varepsilon)\,(\alpha-\beta)^2\,(Q_{LT}(t) - Q_{RT}(t))\,(1-2Q_{RT}(t))\,L(t) \\ \lambda_2 - (\lambda_1+\lambda_2)\,Q_{LT}(t) - \frac{1}{2}\,g^2(\varepsilon)\,Q_{LT}(t)\,(1-Q_{LT}(t))\,(1-2Q_{LT}(t)) \\ \lambda_4 - (\lambda_3+\lambda_4)\,Q_{RT}(t) - \frac{1}{2}\,g^2(\varepsilon)\,Q_{RT}(t)\,(1-Q_{RT}(t))\,(1-2Q_{RT}(t)) \end{bmatrix} .$$

Once again, the increment of time, Δt, was small (at most 10 percent) compared to the mean time between the fastest transitions of the Markov Chain.

The initial values of L, Q_{LT} and Q_{RT} were chosen according to Eq. 6-30. Once a value was obtained for L(t) on each interval, that value was used by

another routine which performed a prediction calculation using Eq. 6-34 and a "measurement" update using Eq. 6-37. Both of these calculations involve only straightforward computations.

IV. FRONT-END/BACK-END PROCESSOR WITH AVERAGING ASSUMPTION

The simulation of this filter is very straightforward, using Eqs. 6-46 and 6-47 to obtain L(k) on each interval. A fast routine simply calculates the integral of dy(t) by performing a summation of the form:

$$\bar{y} = \sum_{kT(\varepsilon)<\tau<(k+1)T(\varepsilon)} dy(\tau) \quad . \tag{8-25}$$

The "slow" routine was then used to propagate the probability p_L in a two-step fashion with a prediction step and an update step as in III above.

8.2.2 Simulation in a Noiseless Environment

In general, simulation of the filtering equations for the noiseless environment is simpler than in the noisy environment because the equations are no longer driven by a Wiener process.

I. EXACT PROBABILITY EQUATIONS

The exact equations for the probabilities of being in the left or right states are given by Eq. 6-49. Equations 6-49a and 6-49b are ordinary differential equations which can be solved using a simple integration routine. In this case the Runge-Kutta technique requires the calculation of

$$\bar{p}_L(t+\Delta t) = p_L(t) + \dot{p}_L(t) \cdot \Delta t \tag{8-26a}$$

$$p_L(t+\Delta t) = p_L(t) + \frac{1}{2}\left[\dot{p}_L(t) + \dot{\bar{p}}_L(t+\Delta t)\right] \Delta t \quad . \tag{8-26b}$$

These steps were implemented with time intervals such that Δt was small compared to the time between the fastest transitions in the Markov chain. The calculations were performed between jump times with an additional integration for the residual time $\tilde{\Delta}t_N$ just preceding each jump:

$$\tilde{\Delta}t_N = (t_N - t_{N-1}) - \left\lfloor \frac{t_N - t_{N-1}}{\Delta t} \right\rfloor \Delta t \qquad (8\text{-}27)$$

and

$\lfloor x \rfloor$ denotes the largest integer $\leq x$.

At jump times, Eqs. 6-49c - 6-49d were used to update the probabilities.

II. FRONT-END/BACK-END PROCESSOR WITH ΔK, ΔR_t

In this case, the processor structure in which ΔR_t and ΔK, the increments in the upper state residence time and the total number of counts respectively are determined for individual segments of time by a fast processor. These values are then used by a slow processor to calculate state probabilities.

The front-end processor determines two quantities, R(t) and K(t) where

$$R(t_N) = R(t_{N-1}) + (t_N - t_{N-1})\, x^{+}(t_{N-1}) \qquad (8\text{-}28a)$$

$$K(t_N) = K(t_{N-1}) + |x(t_N) - x(t_{N-1})| \qquad (8\text{-}28b)$$

where

$$x(t_K) = \begin{cases} 0 & \text{if } \rho(t) \text{ is in state 2 or 4 at time } t_K. \\ 1 & \text{if } \rho(t) \text{ is in state 1 or 3 at time } t_K. \end{cases}$$

Once these quantities are obtained for a time interval $[kT(\epsilon), (k+1)T(\epsilon))$, they are used by the slow processor to update probabilities using the

differential equation given by Eq. 6-56. In this case we must be careful when integrating the equation because of the dK(t) driving term which represents increments in a jump process. The equations that we would normally implement are in the standard Runge-Kutta format

$$\bar{p}_L(t+\Delta t) = p_L(t) + F_1(p_L(t))\Delta R_t + F_2(p_L(t))\Delta t + G(p_L(t))\Delta D + G_2(p_L(t)\Delta U \tag{8-29}$$

using the terminology of Eq. 6-59 to rewrite Eq. 6-56. The second step of the integration is performed using

$$\begin{aligned} p_L(t+\Delta t) = p_L(t) &+ \left[F_1(p_L(t) + F_1(\bar{p}_L(t+\Delta t)\right] \frac{\Delta R_t}{2} \\ &+ \left[F_2(p_L(t)) + F_2(\bar{p}_L(t+\Delta t))\right] \frac{\Delta t}{2} \\ &+ \left[G_1(p_L(t)) + G_1(\bar{p}_L(t+\Delta t))\right] \frac{\Delta D}{2} \\ &+ \left[G_2(p_L(t)) + G_2(\bar{p}_L(t+\Delta t))\right] \frac{\Delta U}{2} \end{aligned} \tag{8-30}$$

Let us analyze the error introduced by the use of Eqs. 8-29 and 8-30. First, if we perform Taylor series expansions of $F_1(\bar{p}_L(t+\Delta t))$, etc., in Eq. 8-30 and keep only terms of zeroth and first order, we have

$$\begin{aligned} p_L(t+\Delta t) = p_L(t) &+ F_1(p_L(t))\Delta R_t + F_2(p_L(t))\Delta t + G_1(p_L(t))\Delta D + G_2(p_L(t))\Delta U \\ &+ \frac{1}{2}\left[F_1'(p_L(t))\Delta R_t + F_2'(p_L(t))\Delta t + G_1'(p_L(t))\Delta D + G_2'(p_L(t))\Delta U\right] \cdot \\ &\quad \left[F_1(p_L(t))\Delta R_t + F_2(p_L(t))\Delta t + G_1(p_L(t))\Delta D + G_2(p_L(t))\Delta U\right] \cdot \end{aligned} \tag{8-31}$$

Now, recalling the definition of the functions F_1, F_2, G_1 and G_2,

$$F_1(p_L(t)) = \varepsilon[\mu_2-\mu_4 - (\mu_1+\mu_2-\mu_3-\mu_4)p_L(t)] + (\alpha-\beta)g(\varepsilon)p_L(t)(1-p_L(t)) \tag{8-32a}$$

$$F_2(p_L(t)) = \varepsilon[\mu_4 - (\mu_3+\mu_4)p_L(t)] + \beta\, g(\varepsilon)\, p_L(t)\, (1-p_L(t)) \tag{8-32b}$$

$$G_1(p_L(t)) = \frac{-g(\varepsilon)\, \alpha p_L(t)\, (1-p_L(t))}{\lambda_1 + \alpha g(\varepsilon)\, (1-p_L(t))} \tag{8-32c}$$

$$G_2(p_L(t)) = \frac{-g(\varepsilon)\, \beta p_L(t)\, (1-p_L(t))}{\lambda_2 + \beta g(\varepsilon)\, (1-p_L(t))} \tag{8-32d}$$

and substituting into Eq. 8-31 we obtain

$$\begin{aligned} p_L(t+\Delta t) \cong\ & p_L(t) + \left[(\alpha-\beta)g(\varepsilon)\, p_L(1-p_L) + \varepsilon[(\mu_2-\mu_4) - (\mu_1+\mu_2-\mu_3-\mu_4)p_L(t)]\right]\Delta R_t \\ & + \left[\beta g(\varepsilon)\, p_L(1-p_L) + \varepsilon[\mu_4 - (\mu_3+\mu_4)p_L(t)]\right]\Delta t \\ & - \frac{\alpha g(\varepsilon)\, p_L(t)\, (1-p_L(t))}{\lambda_1 + \alpha g(\varepsilon)\, (1-p_L(t))}\, \Delta D - \frac{\beta g(\varepsilon)\, p_L(t)\, (1-p_L(t))}{\lambda_2 + \beta g(\varepsilon)\, (1-p_L(t))}\, \Delta U \\ & + p_L(t)\, (1-p_L(t))\, (1-2p_L(t))\, \frac{g^2(\varepsilon)}{2} \cdot \left\{ \left[(\alpha-\beta)^2\, \Delta R_t^2 + \beta^2\, \Delta t^2 + \frac{\alpha^2}{\lambda_1^2}\, \Delta D^2 + \frac{\beta^2}{\lambda_2^2}\, \Delta U^2 \right] \right. \\ & + 2 \left[\beta(\alpha-\beta)\Delta t\, \Delta R_t - \frac{\alpha(\alpha-\beta)}{\lambda_1}\, \Delta R_t\, \Delta D - \frac{\beta(\alpha-\beta)}{\lambda_2}\, \Delta R_t\, \Delta U - \frac{\alpha\beta}{\lambda_1}\, \Delta t\, \Delta D - \frac{\beta^2}{\lambda_2}\, \Delta t\, \Delta U \right. \\ & \left. \left. + \frac{\alpha\beta}{\lambda_1\lambda_2}\, \Delta D\, \Delta U \right] \right\} \end{aligned} \tag{8-33}$$

To compare Eq. 8-33 to the exact equations, consider the relationship between Δp_L and ΔL for small ΔL. Substituting Eq. 8-34 into Eq. 6-65, we obtain Eq. 8-35 and finally Eq. 8-36

$$L(t+\Delta t) = L(t) + \Delta L \quad . \tag{8-34}$$

$$\Delta L = \frac{-\Delta p_L(t)}{p_L(t)\,(1-p_L(t))} + O((\Delta L)^2) \tag{8-35}$$

$$\Delta L = \left[-\beta\Delta t - (\alpha-\beta)\Delta R_t + \frac{\alpha}{\lambda_1}\Delta D + \frac{\beta}{\lambda_2}\Delta U \right] g(\varepsilon) + O(g^2(\varepsilon)) \Bigg] \quad . \tag{8-36}$$

Expanding the exact expression for L, (Eq. 6-63), using a Taylor expansion for the exponential and Binomial theorem for the power factors, we obtain exactly Eq. 8-36.

Therefore we see that the basic Runge-Kutta technique will be accurate to within $O(g^2(\varepsilon))$.

III. FRONT-END/BACK-END PROCESSOR USING LIKELIHOOD FUNCTION

Simulations were performed for the front-end/back-end filter structures for which L(t), the likelihood ratio described by Eq. 6-64 is calculated using the equation given by Eq. 6-63. The calculation of L(t) is straightforward and performed over time increments of length $T(\varepsilon)$, based upon the ΔR_t, ΔD, ΔU statistics determined by the fast integrator and counters of the "front end." These statistics were determined from the sample path in the same manner as described for II above. The values of L(t) were calculated and passed to a "slow routine" which calculated the state probabilities using a two-stage process comprised of prediction and update calculations.

The prediction step was implemented using the "approximation technique" given by Eqs. 6-76 and 6-77. For the choice of μ's that was used, however (μ_i=1), the condition of Eq. 6-74 was satisfied so that Eqs. 2-76 and 2-77 were exact. The update calculation was done using Eq. 6-78.

8.3 SIMULATION RESULTS

8.3.1 Probability Propagation for White Noise Model

In subsection 8.2, the simulations for four filtering techniques in the noisy observation case were described. The results for each of these cases are provided in Figs. 8-1 through 8-14 and are discussed below. The plots of the filter outputs show two pieces of information. The "square wave" plot shows the sample path of the system in terms of left-to-right transitions. A high value in the plot (.75) indicates the system is actually in the left pair of states and a low value (.25) indicates the system is actually on the right. The probability of being in the left pair of states as determined by the filter is superimposed on this plot. Since a large value of p_L represents a conclusion that we are on the left and a small p_L that we are on the right, an indication of filter performance is the amount of time that the filter output and the sample path are both in the top or bottom of the plot. This provides a qualitative measure of how often the filter reaches the correct "conclusion."

I. EXACT FILTER

The simulations for this and all other white noise cases used the parameter values given in Eq. 8-37

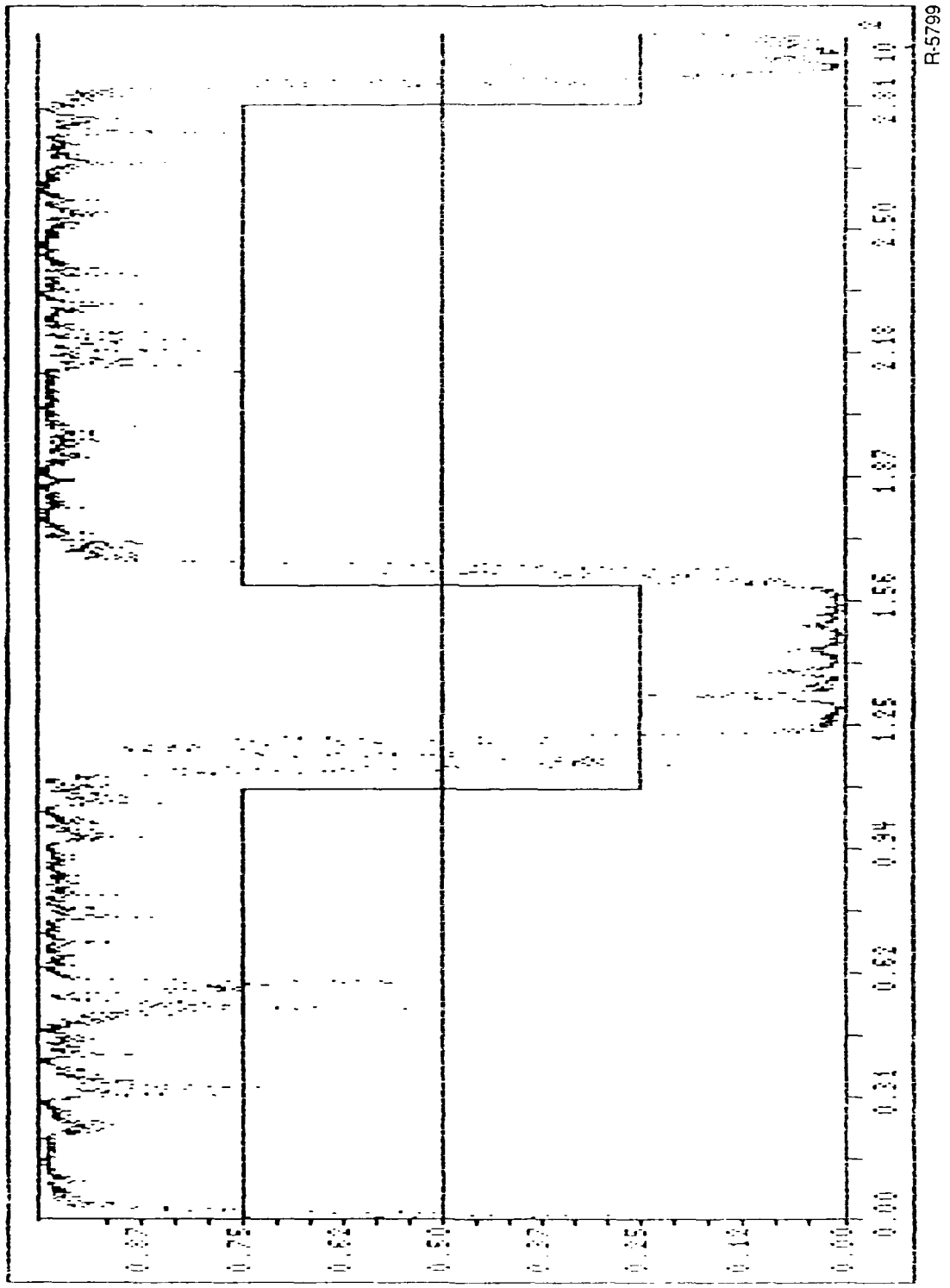


Figure 8-1. Exact Probabilities for White Noise Model, $g = 1$, $\epsilon = .01$, $\Delta t = .1$

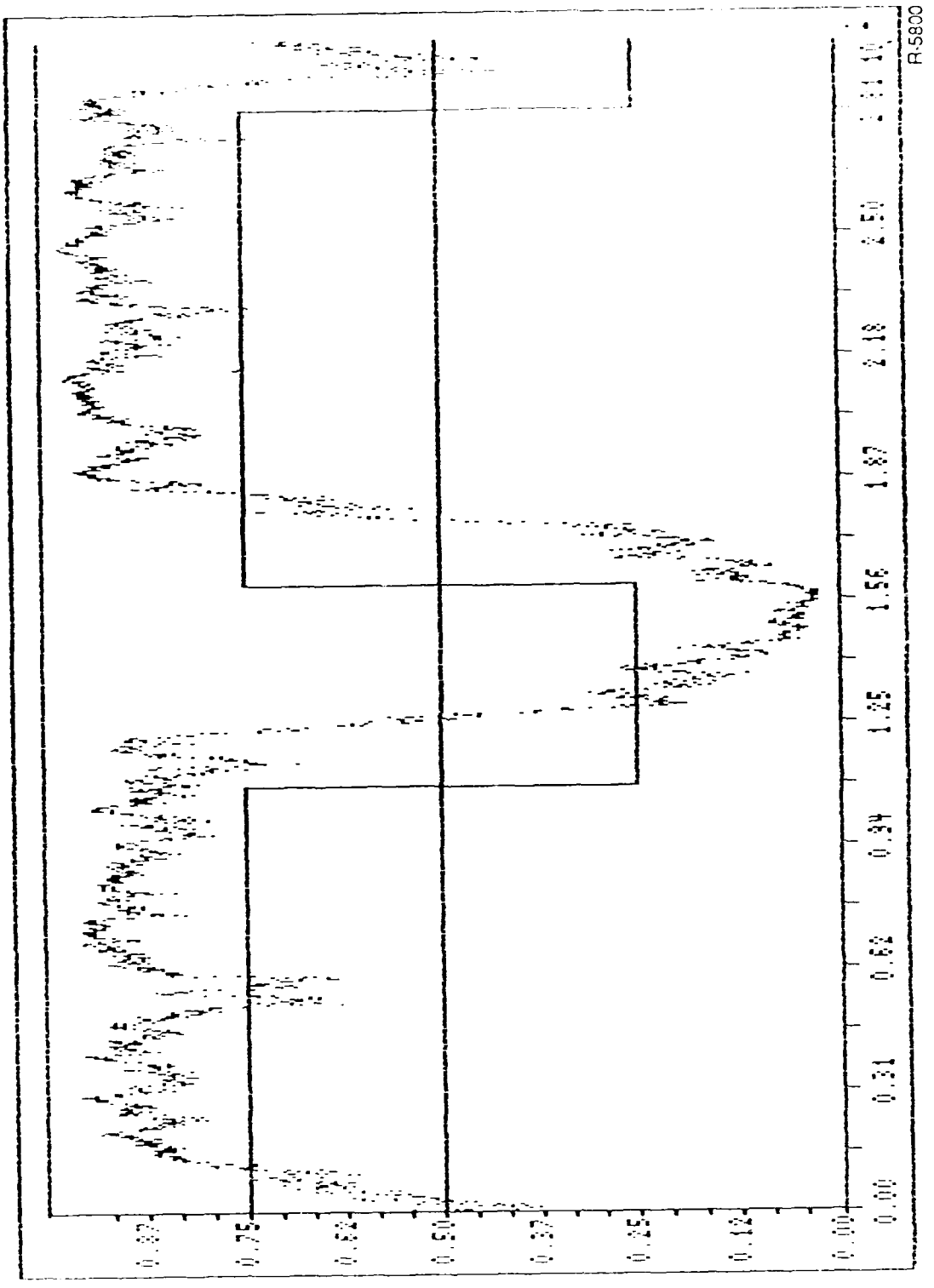


Figure 8-2. Exact Probabilities for White Noise Model, $g = .3$, $\varepsilon = .01$, $\Delta t = .1$

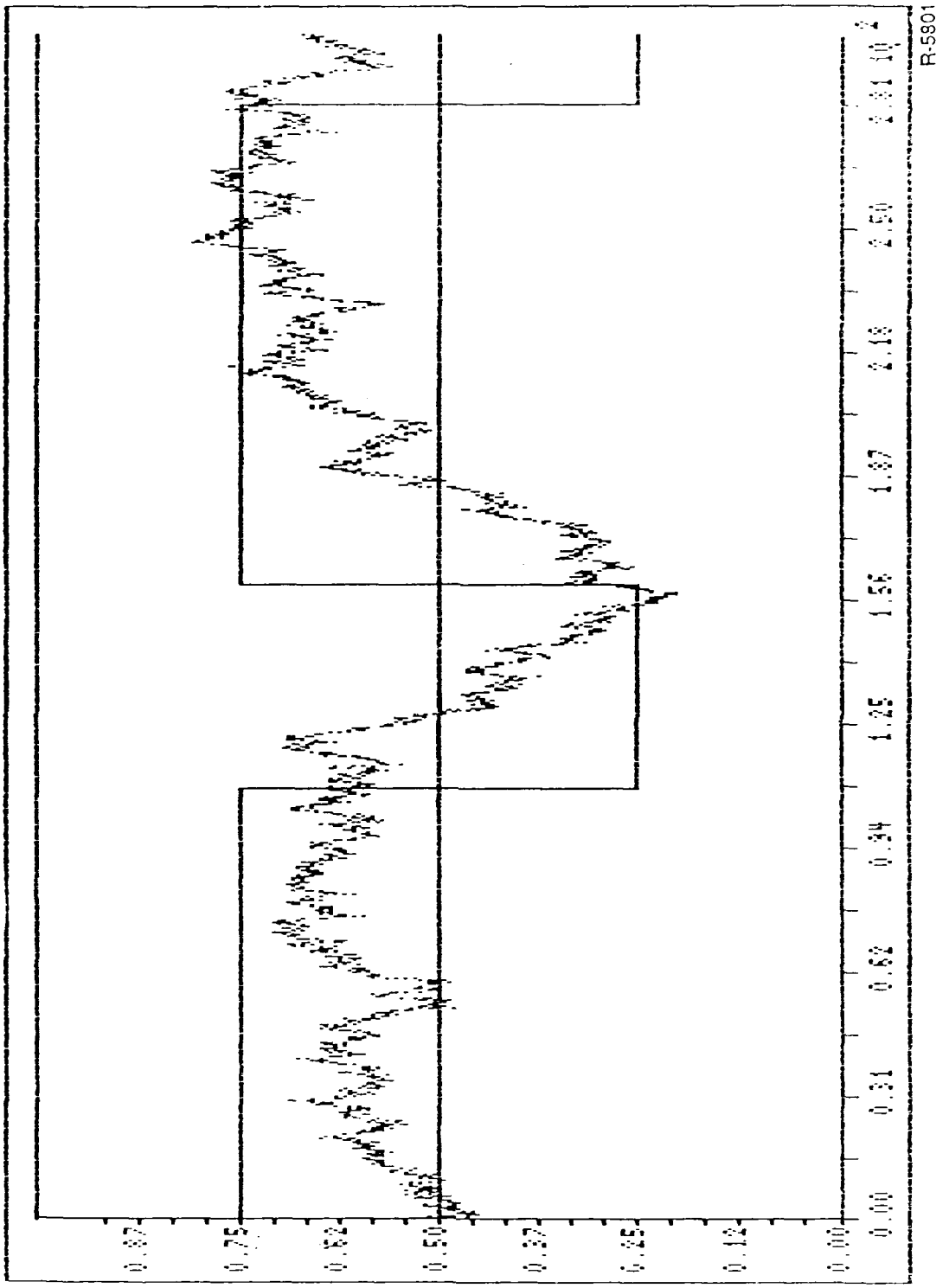


Figure 8-3. Exact Probabilities for White Noise Model, $g = .1$, $\epsilon = .01$, $\Delta t = .1$

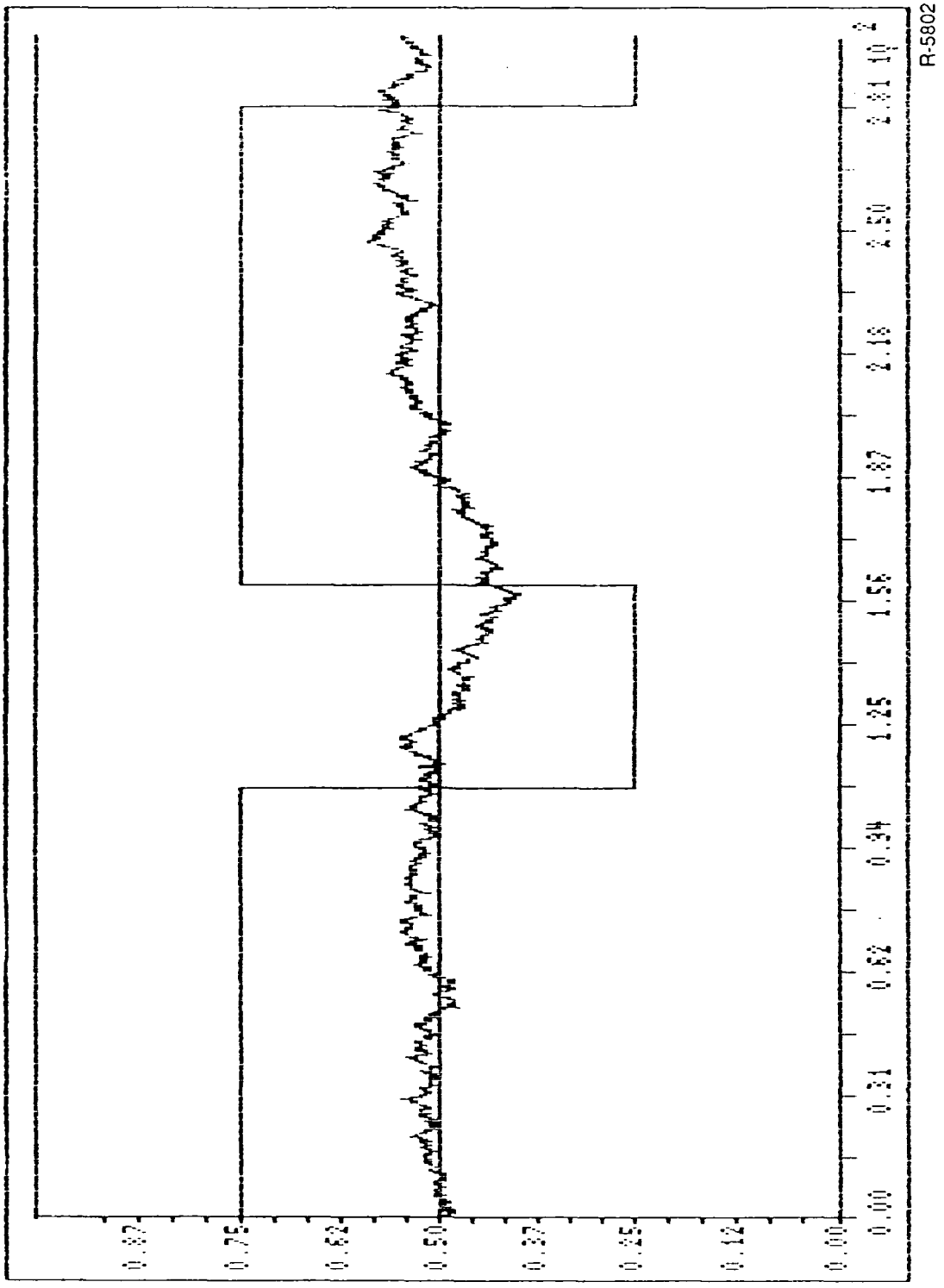


Figure 8-4. Exact Probabilities for White Noise Model, $g = .03$, $\epsilon = .01$, $\Delta t = .1$

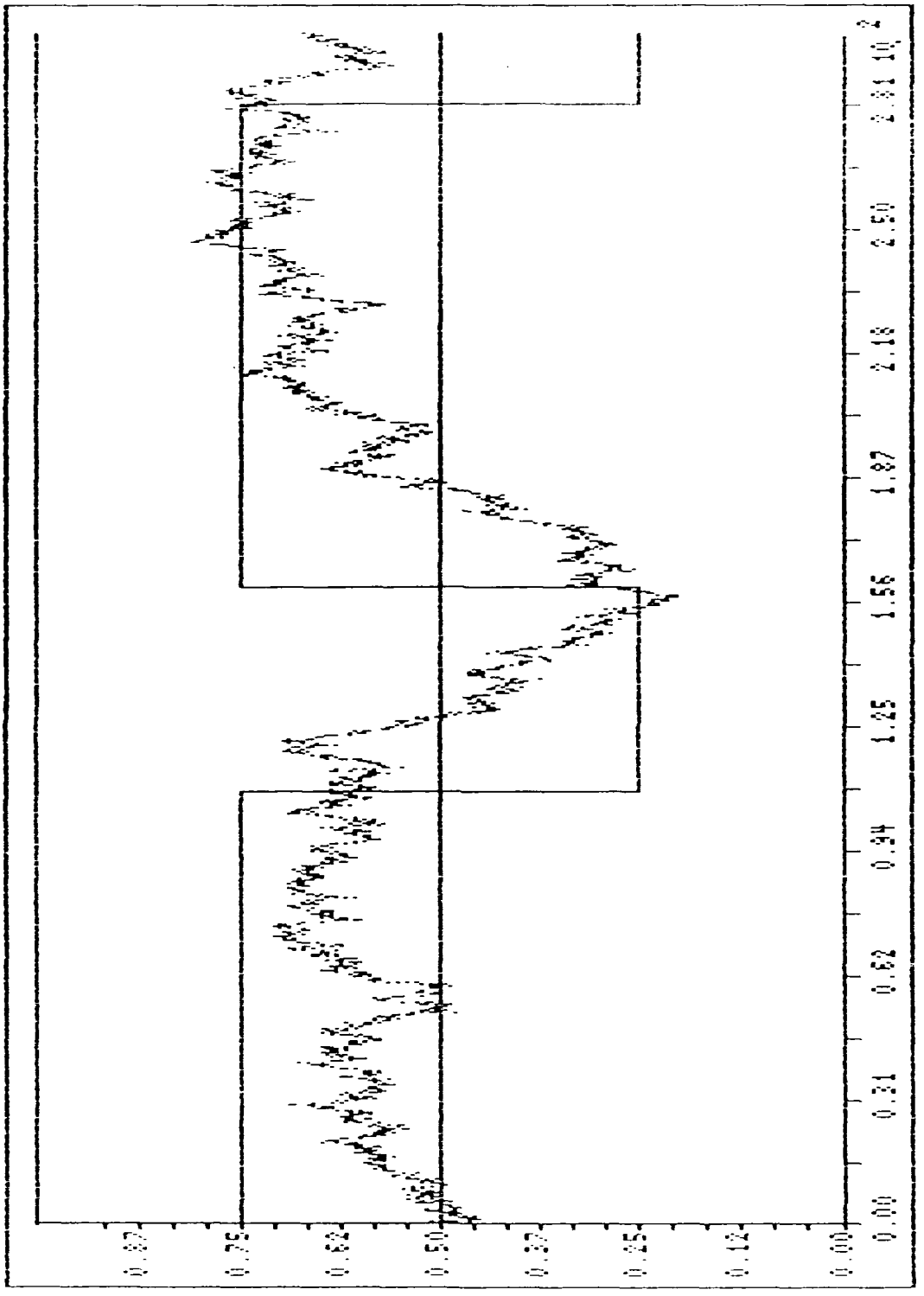


Figure 8-5. Probabilities for Aggregate White Noise Model, $\beta = .1$, $\epsilon = .01$, $\Delta t = 0.1$

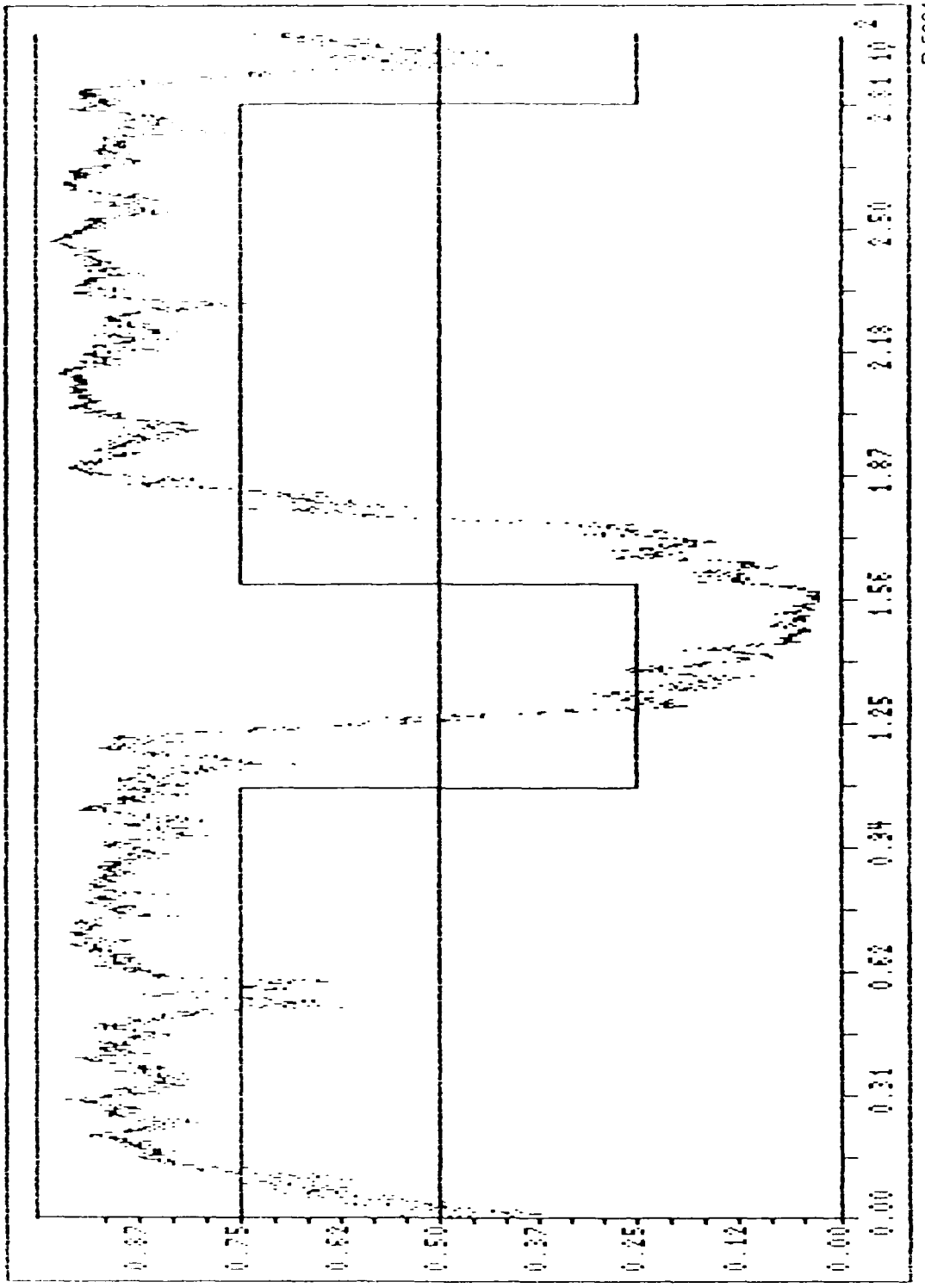


Figure 8-6. Probabilities for Aggregate White Noise Model, g = .03, ε = 0.01, Δt = 0.1

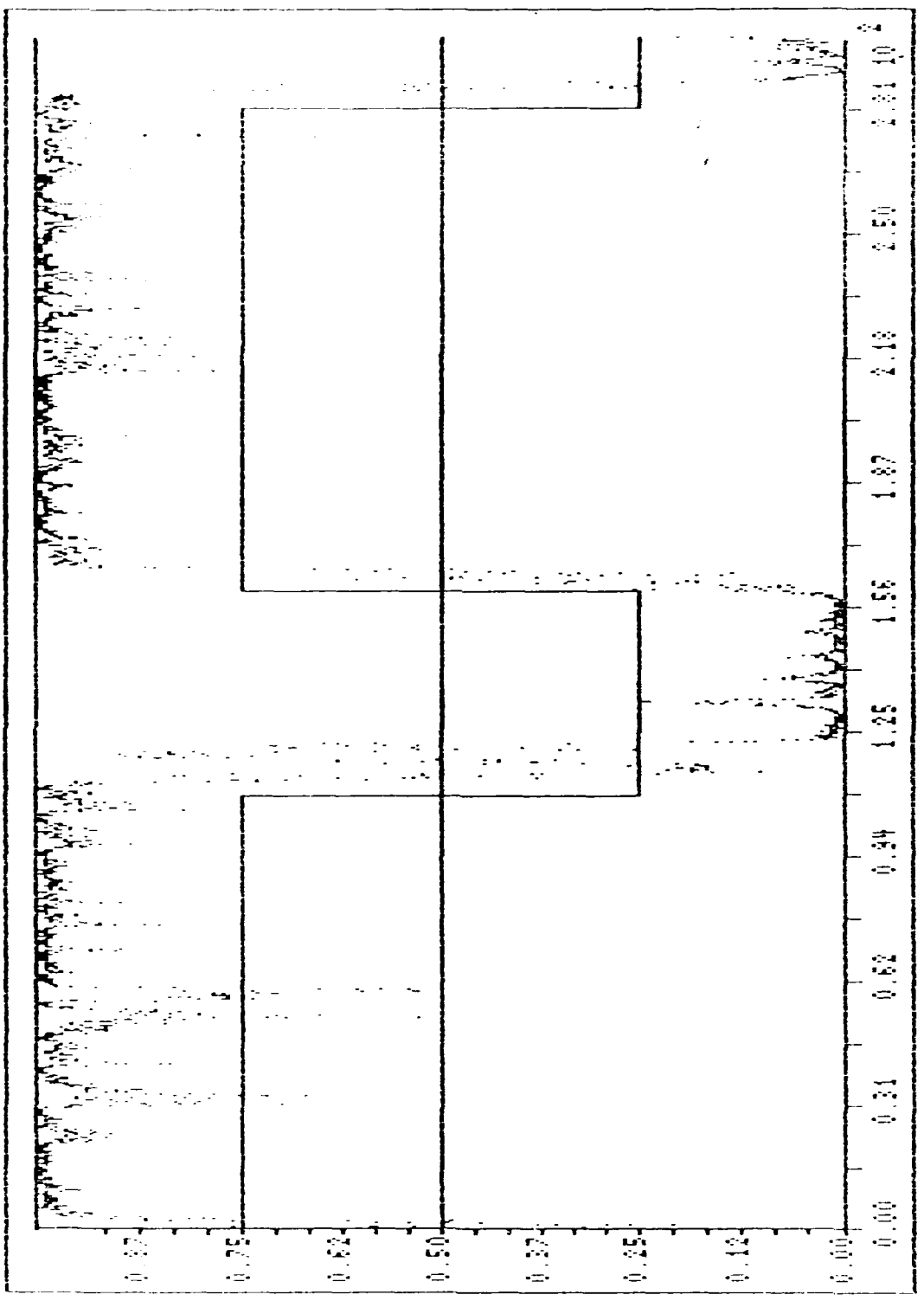


Figure 8-7. Probabilities for Aggregate White Noise Model, $g = 1$, $\epsilon = .01$, $\Delta t = 0.1$

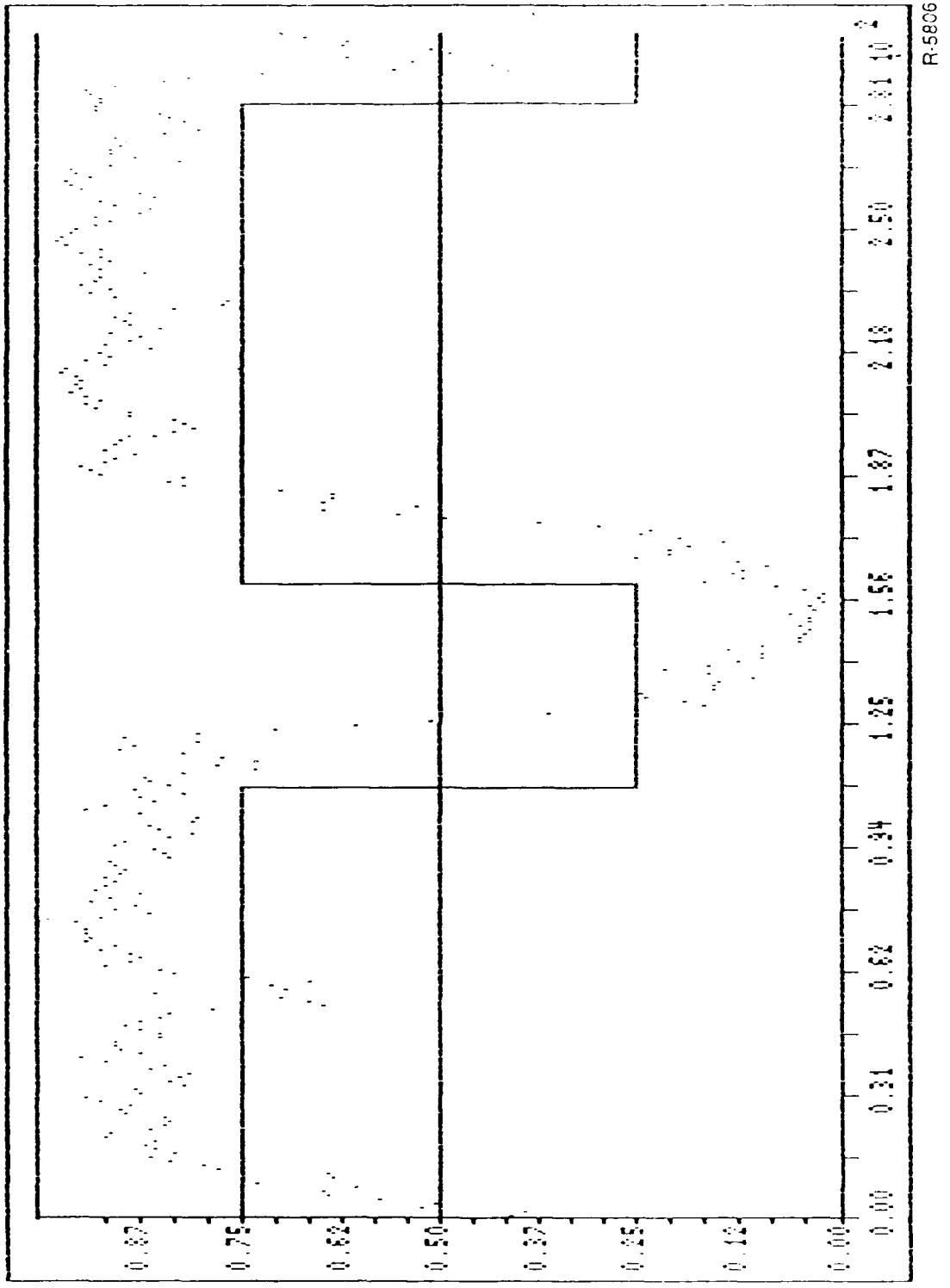


Figure 8-8. FE/BE Simulation for White Noise Model with $g = .3$, $\epsilon = 0.01$, $\Delta t = 1$

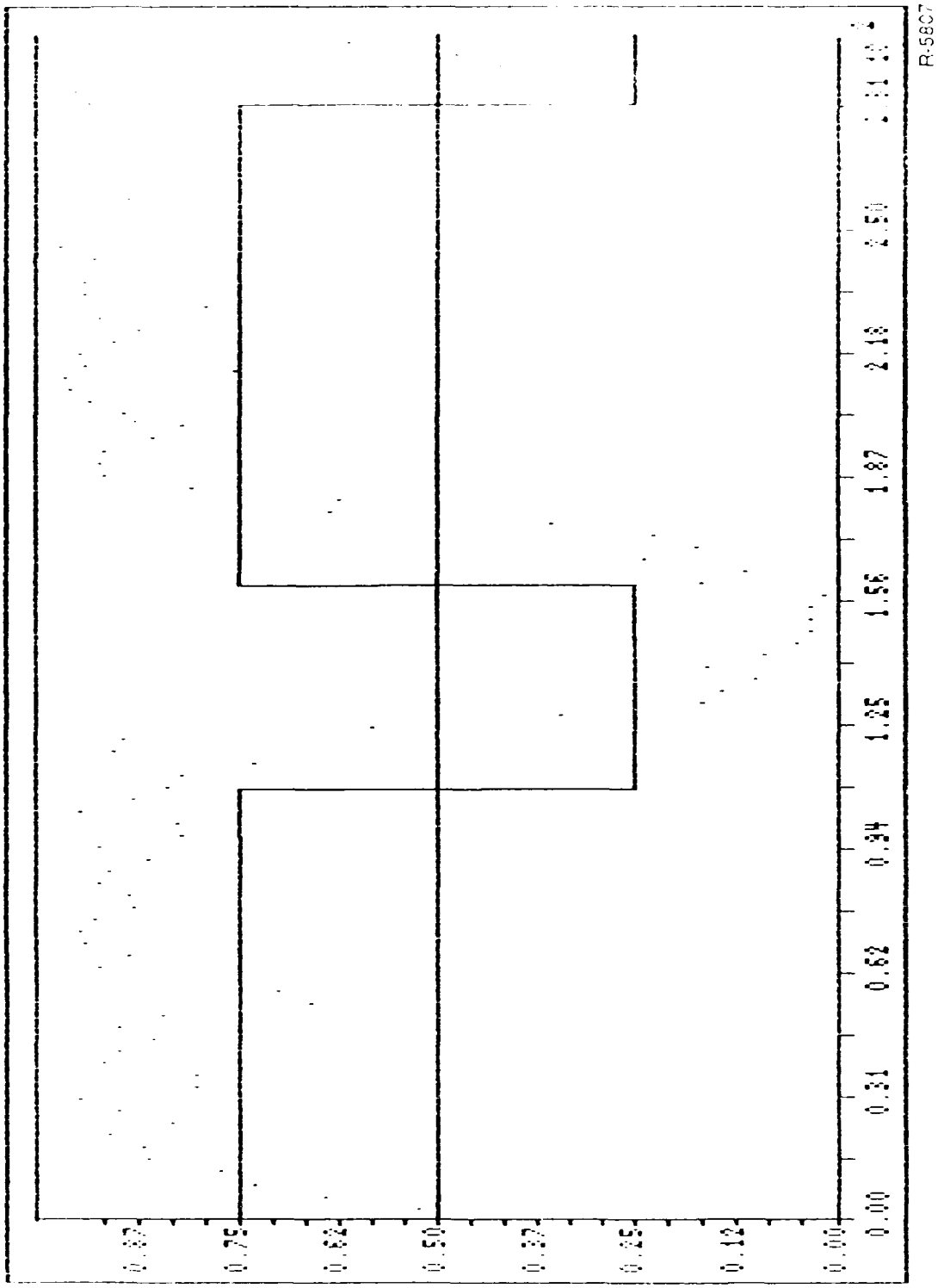


Figure 8-9. FE/BE Simulation for White Noise Model with $g = .3$, $\epsilon = .01$, $\Delta t = 3$

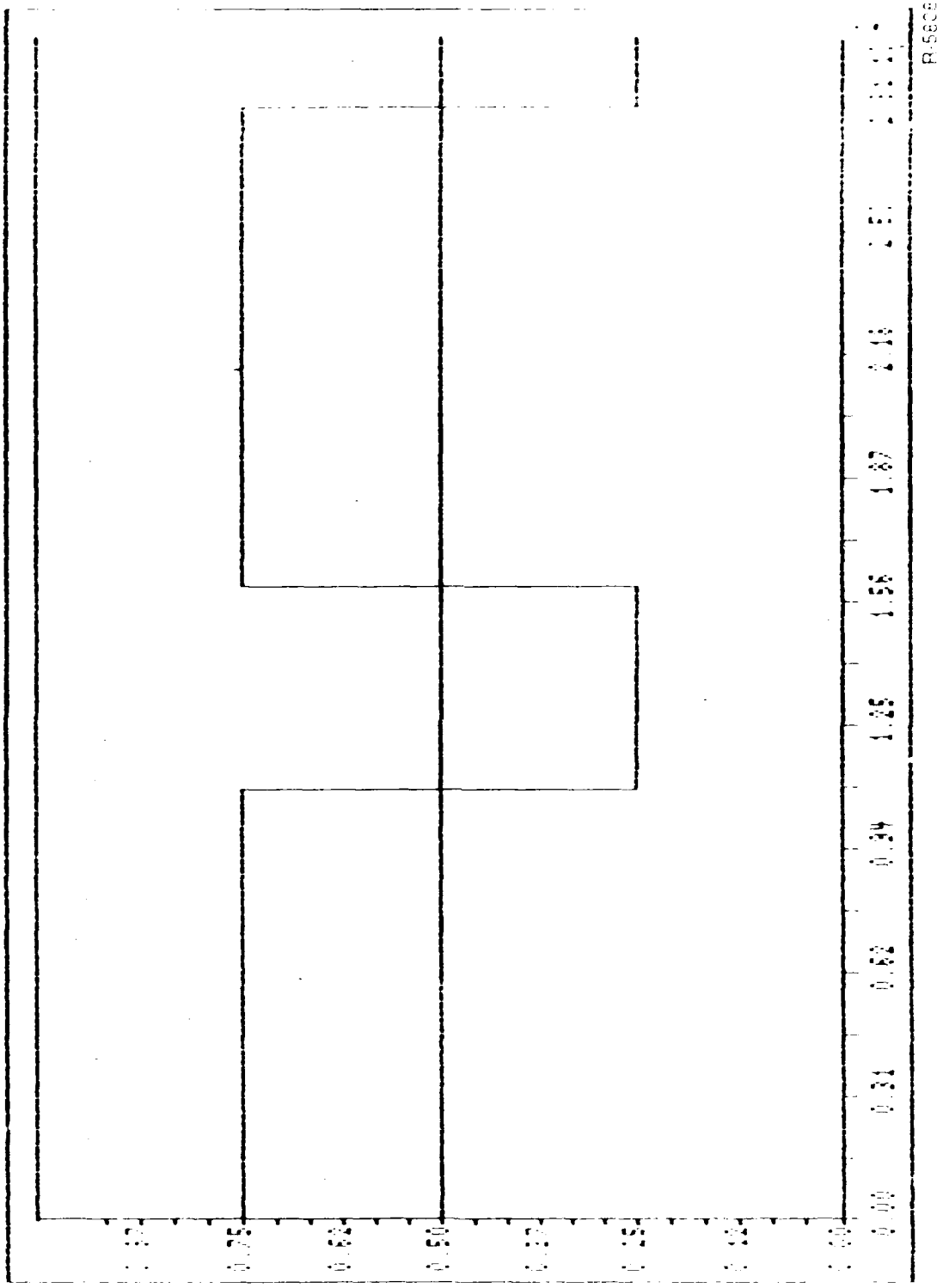


Figure 8-10. FE/BE Simulation for White Noise Model, $\beta = .3$, $\epsilon = .01$, $\Delta t = 10$

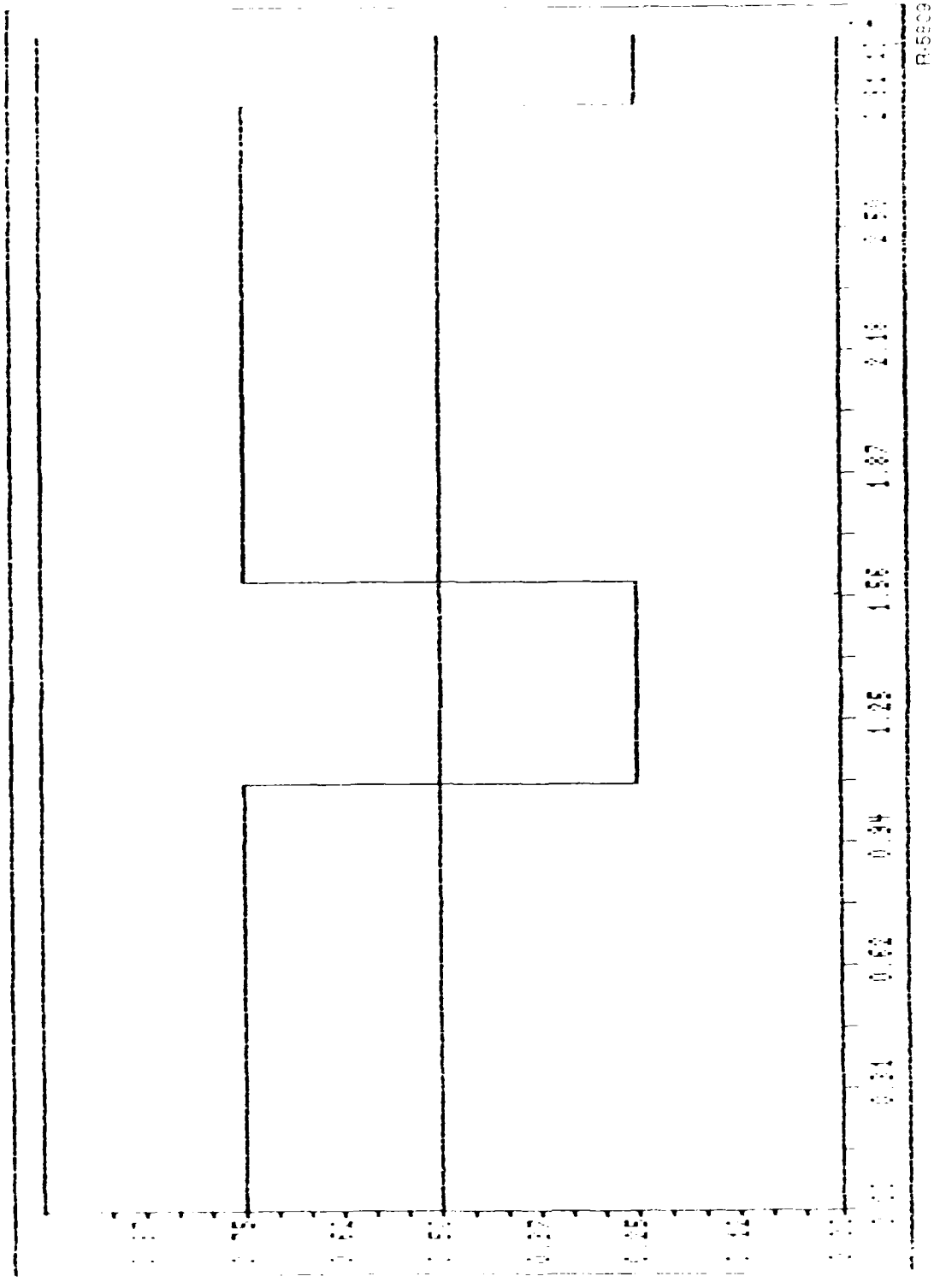


Figure 8-11. FE/BE Simulation for White Noise Model, $g = .3$, $\epsilon = .01$, $\Delta t = 31$

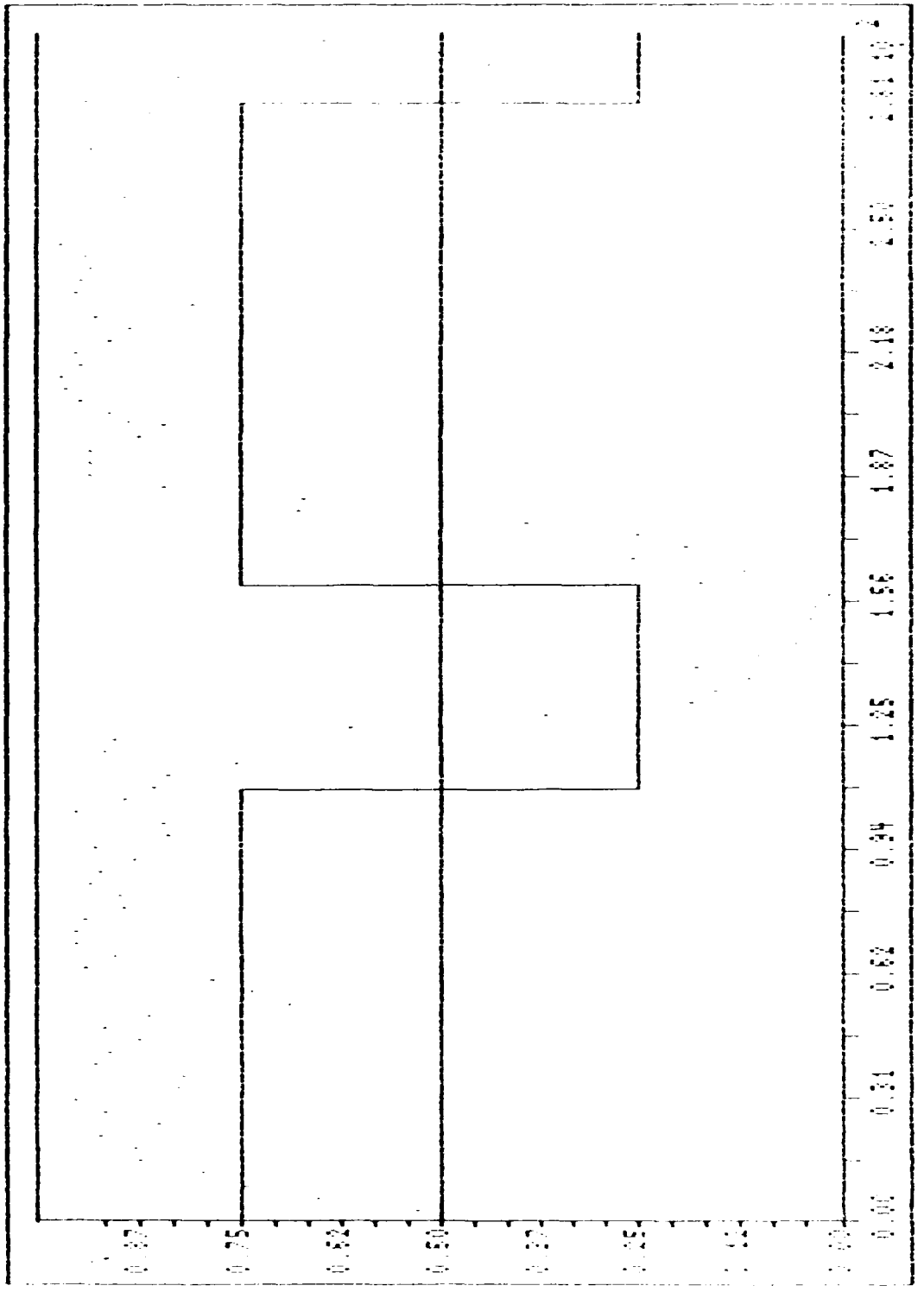


Figure 8-12. FE/BE Simulation for Aggregate and White Noise Model, $\beta = .3$, $\epsilon = .01$, $\Delta t = 3$

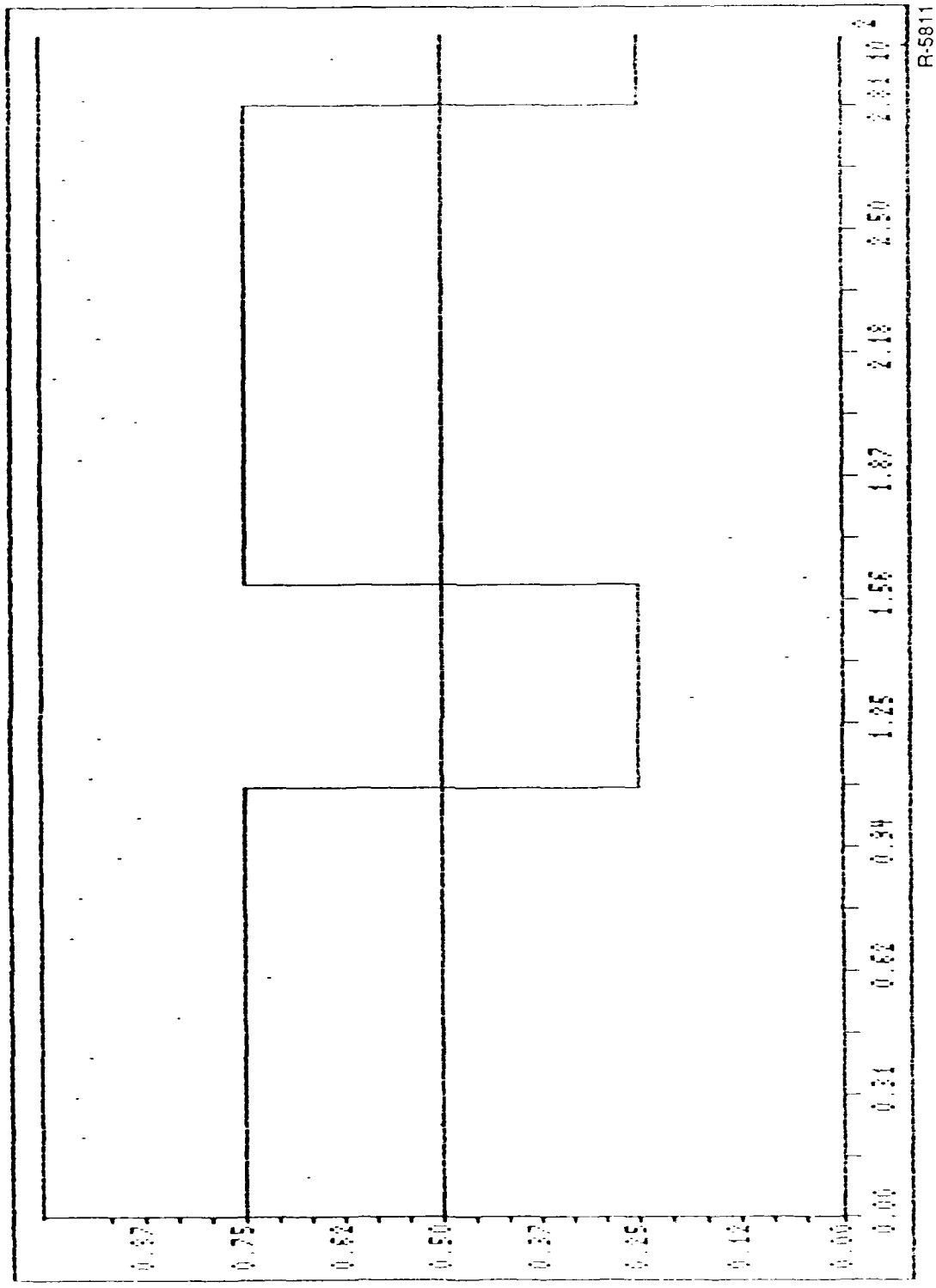


Figure 8-13. FE/BE Simulation for Aggregate and White Noise Model, $g = .3$, $\epsilon = .01$, $\Delta t = 10$

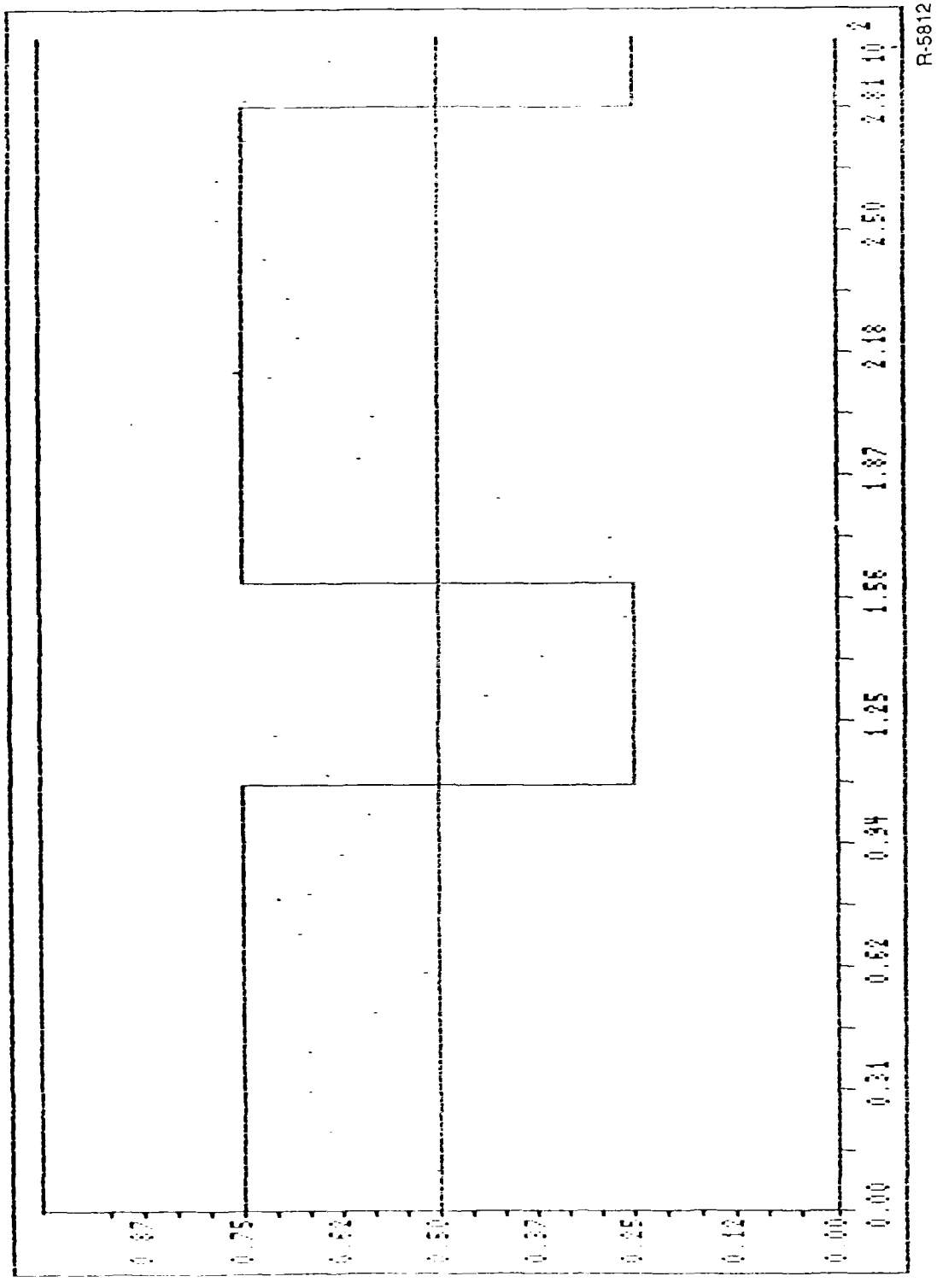


Figure 8-14. FE/BE Simulation for Aggregate and White Noise Model, $g = .1$, $\varepsilon = .01$, $\Delta t = 10$

$$\lambda_1 = \lambda_4 = 1 \tag{8-37a}$$

$$\lambda_2 = \lambda_3 = 5 \tag{8-37b}$$

$$\mu_1 = \mu_2 = \mu_3 = \mu_4 = 1 \tag{8-37c}$$

$$\Delta t = 0.1 \tag{8-37d}$$

$$\varepsilon = 0.01 \tag{8-37e}$$

where Δt is the time increment between calculations.

The results for four simulations are provided for the exact filter case. Each simulation was done with an identical sample path, but with different values for $g(\varepsilon)$ the signal magnitude in the output. With a noise intensity of 1, Figs. 8-1 through 8-4 provide results for $g(\varepsilon) = 1.00$, 0.30, 0.10, and 0.03 corresponding to ε^0, $\varepsilon^{1/4}$, $\varepsilon^{1/2}$, and $\varepsilon^{3/4}$, respectively.

Several features of the results are worth noting. The filter appears to display a "switch" type of behavior for larger values of $g(\varepsilon)$, particularly $g(\varepsilon) = 1$. For $g(\varepsilon) = \varepsilon^{1/4}$, performance deterioriates somewhat, but still provides a correct result in the sense that $p_L > 0.5$ for the majority of time that the system is in the left pair of states and $p_L < 0.5$ when the process is on the right. When the signal strength is decreased to $g(\varepsilon) = \varepsilon^{1/2}$, performance becomes somewhat worse, with smaller excursions away from the unconditional probability value of 0.5 and with occasional excursions to the incorrect side of $p_L = 0.5$, without a transition having occurred. Finally, when $g(\varepsilon) = \varepsilon^{1/2}$, performance breaks down substantially. Although the probabilities move in the same general direction as in the $\varepsilon^{1/2}$ case, excursions from $p_L = 0.5$ are quite small, and bursts of noise occasionally generate the incorrect conclusion that a transition has occurred.

II. AGGREGATE PROBABILITIES FOR WHITE NOISE MODEL

The output of the aggregate filter described in Section 6 is presented next for the case of $\varepsilon = 0.01$, $g(\varepsilon) = 0.1$, 0.3 and 1.0 (Figs. 8-5 through 8-7). The results of Section 7 indicate that for $g(\varepsilon) = O(\varepsilon^{1/2})$, the performance of this filter should be asymptotically close to that of the optimal filter as ε decreases to 0. Comparison of the simulation outputs of the aggregate and exact filters shows almost exact agreement for the cases of $g(\varepsilon) = 0.1$ $(\varepsilon^{1/2})$ and $g(\varepsilon) = 0.3$ $(\varepsilon^{1/4})$. In the case of $g(\varepsilon) = 1$, agreement was still good, but deviations of up to 0.05 in magnitude can be found. In all cases, however, the difference between the aggregate and the exact versions of the filter was of much smaller order than the bound of $O(g^2(\varepsilon))$ derived in Section 7.

III. FE/BE SIMULATION FOR WHITE NOISE MODEL

The FE/BE model refers to the case when a sufficient statistic, L, is calculated in batches by a front end with no apriori information. The back end then applies a Bayesian update procedure using successive values of L. The simulations were done with $g = 0.3$ and $\varepsilon = .01$; a case in which the optimal filter performs well, but does not yet act like a "switch." The time intervals, $T(\varepsilon)$ selected for the collection of data were 1, 3, 10 and 31 corresponding to ε^0, $\varepsilon^{-1/4}$, $\varepsilon^{-1/2}$ and $\varepsilon^{-3/4}$, respectively. The plots for these cases are provided in Figs. 8-8 through 8-11.

The simulations show that for relatively small time intervals, namely $T(\varepsilon) = 1$ and $T(\varepsilon) = 3$, performance is excellent, as the probabilities generated, differ from the optimal values by at most .01 for this sample path. It is our conjecture that for $g(\varepsilon)T(\varepsilon) < 1$, excellent performance can be expected, with deteriorating agreement between the exact filter and this approximation

when $g(\varepsilon)T(\varepsilon)$ exceeds 1. The values of this product for $T(\varepsilon) = 1$ and 3 are both less than 1, and we do in fact see very good performance in this case. When we reach $T(\varepsilon) = 10$, we have that $g(\varepsilon)T(\varepsilon) = 3 > 1$ and therefore expect that deviations between the FE/BE filter and the optimal filter will start to increase. This is indeed the case; comparing the FE/BE with ten time unit batches to the exact filter for this sample path shows a maximum deviation of .04 while for $T(\varepsilon) = 31$, in which case $g(\varepsilon)T(\varepsilon) = 9.3$, we have a maximum deviation of .08, with values consistently biased towards the 0 or 1 limits. It is worth noting that although the probabilities generated by the filter differ slightly, the final conclusions regarding the current state of the system seldom differ.

IV. FE/BE SIMULATION FOR AGGREGATE WHITE NOISE MODEL

The final filter approximation, actually a combination of II and III, was simulated with the same parameter values as III, $\varepsilon = .01$, $g(\varepsilon) = .3$. The batch times, $T(\varepsilon)$, were selected to be 3 and 10 (Fig. 8-13). The results are very similar to those in case III, showing excellent agreement with the exact filter for $T(\varepsilon)g(\varepsilon) = 1$ (within .01) and somewhat larger differences (maximum .06) for $T(\varepsilon) = 10$ which yields $g(\varepsilon)T(\varepsilon) = 3$. A final run is provided in Fig. 8-14 in which $g(\varepsilon)$ was decreases to .1 while $T(\varepsilon)$ remained at 10. In this case, $g(\varepsilon)T(\varepsilon) = 1$ so we again expect reasonable agreement with the optimal. Comparing Figs. 8-14 and 8-3 we find that this is indeed the case.

8.3.2 Quantitative Aspects of White Noise Model Simulations

In addition to the simulations based on simple sample paths for which plots were generated, multiple sample paths for individual parameter values were used to estimate numerical characteristics of both the exact filter and

the aggregate approximation. In particular, the quantities of interest were measures of the rate at which information is supplied to the filter by the measurements and in the case of the approximate filter, the difference between the approximate output and the optimal output. The first quantity, the information, is defined in Section 7 as $\mu(t)$ and is the magnitude of the difference between the filter output and the output that would appear in steady-state if no measurements were available. For our four-state model, if we calculate the probability of being in the left pair of states ($p^L(t)$) then

$$\mu(t) = \left| p^L(t) - \frac{\gamma_2}{\gamma_1+\gamma_2} \right| . \qquad (8\text{-}38)$$

In Eq. 8-38, γ_1 and γ_2 are the left-to-right and right-to-left transition rates assuming the fast transitions have reached steady-state and are given by equations defined in subsection 6.2.1. For our numerical example, $\gamma_1 = \gamma_2 = 1$. The measures of these quantities that we are interested in are their suprema and their mean square values.

Parameter Values:

In all simulation runs, the parameter values that were selected are given by Eq. 8-37, and each sample paths had a length of 1000 time units.

Mean Square Results

The first three plots (Figs. 8-15 through 8-17) present the results for the mean square error of the filter based on the aggregate model and the information available in the "exact model." The first graph provides a log-log plot of the mean square value of the information versus the order of ε in the signal power expression. These values are given by Eqs. 8-39 and 8-40.

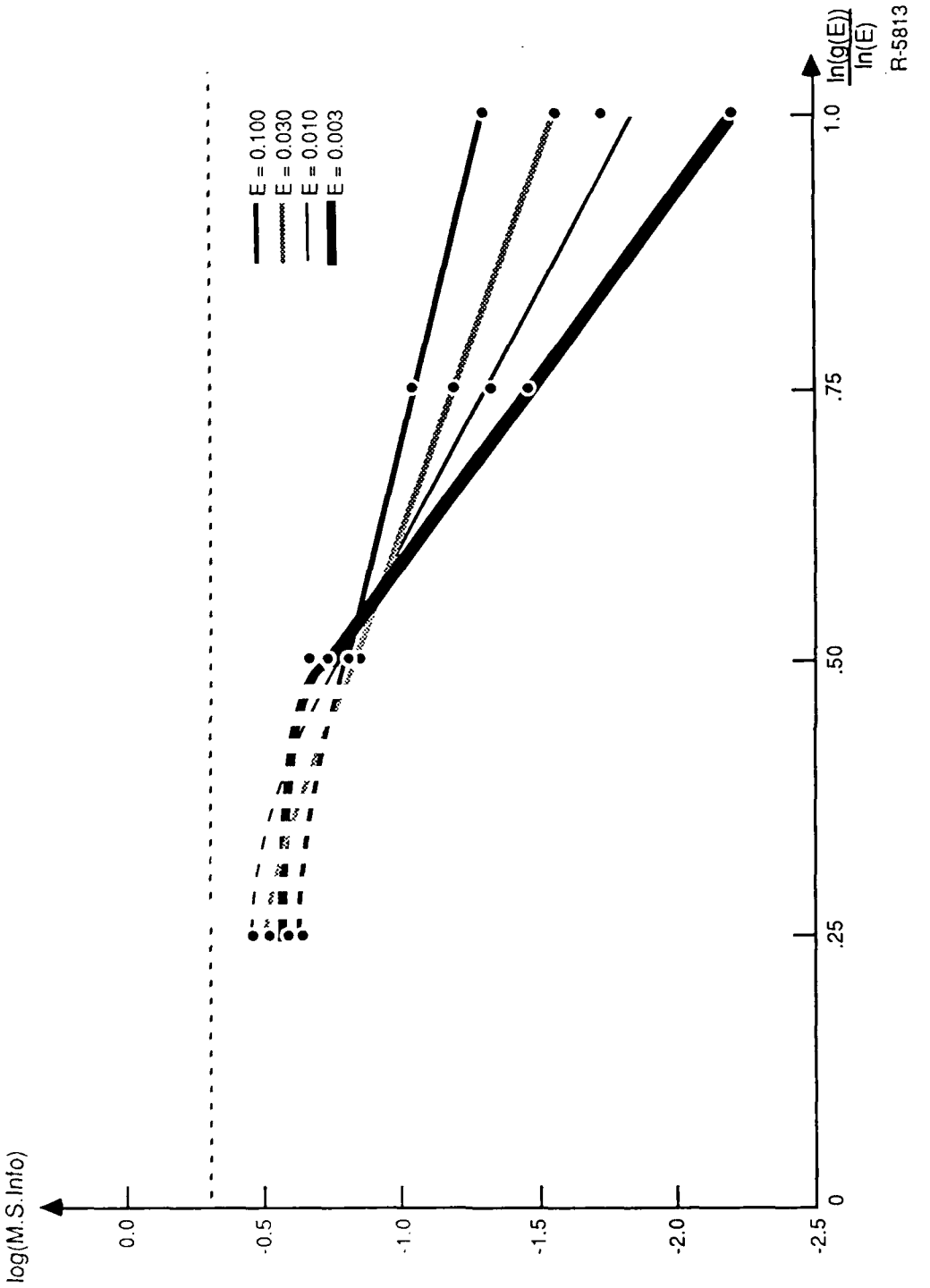


Figure 8-15. Plot of Log $\left[\text{M.S.} \left(p^L - \left(\frac{\gamma_2}{\gamma_1+\gamma_2} \right) \right) \right]$ of Exact Filter

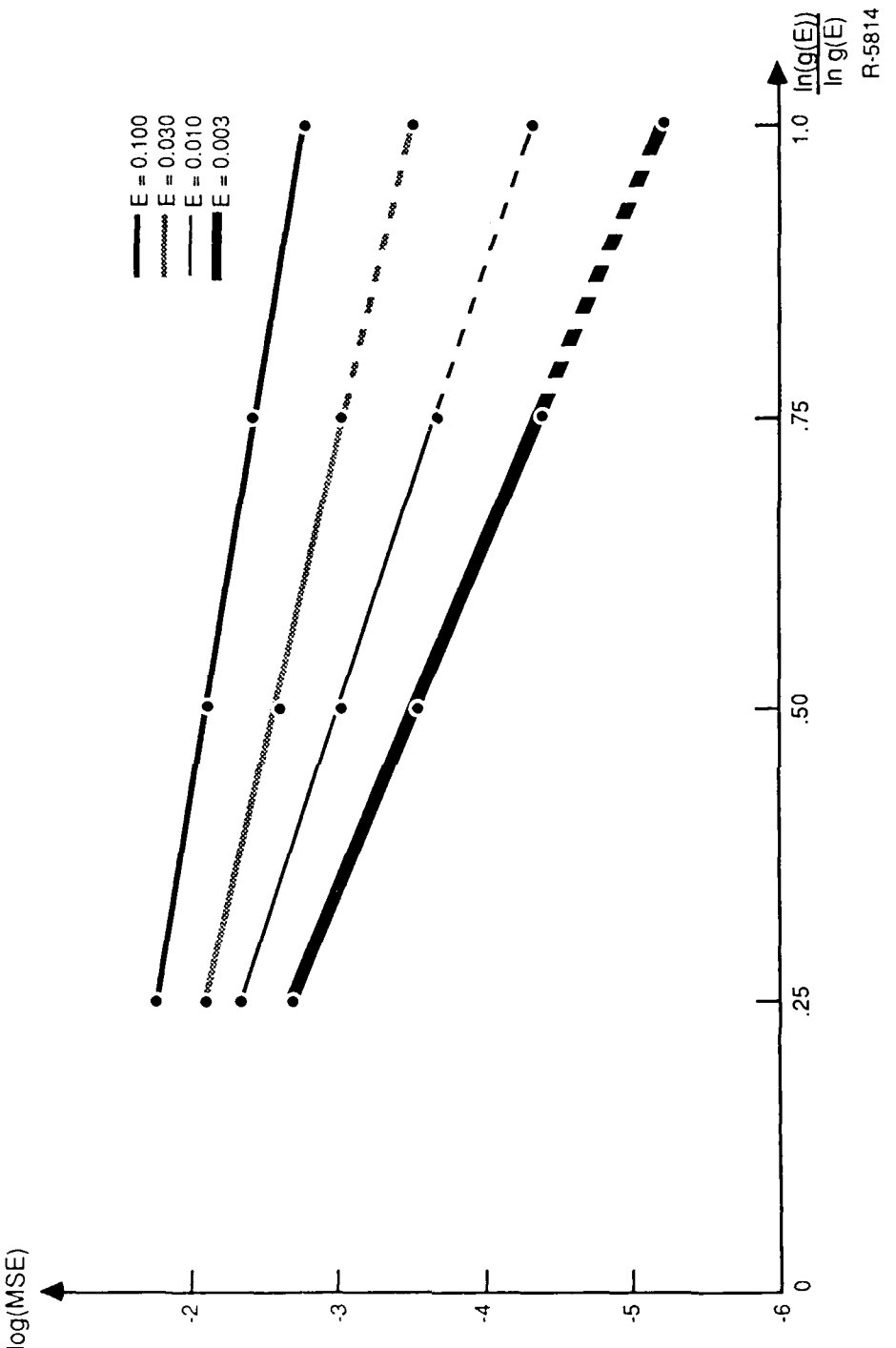


Figure 8-16. Plots of Log (MSE) of Aggregate Filter in White Noise

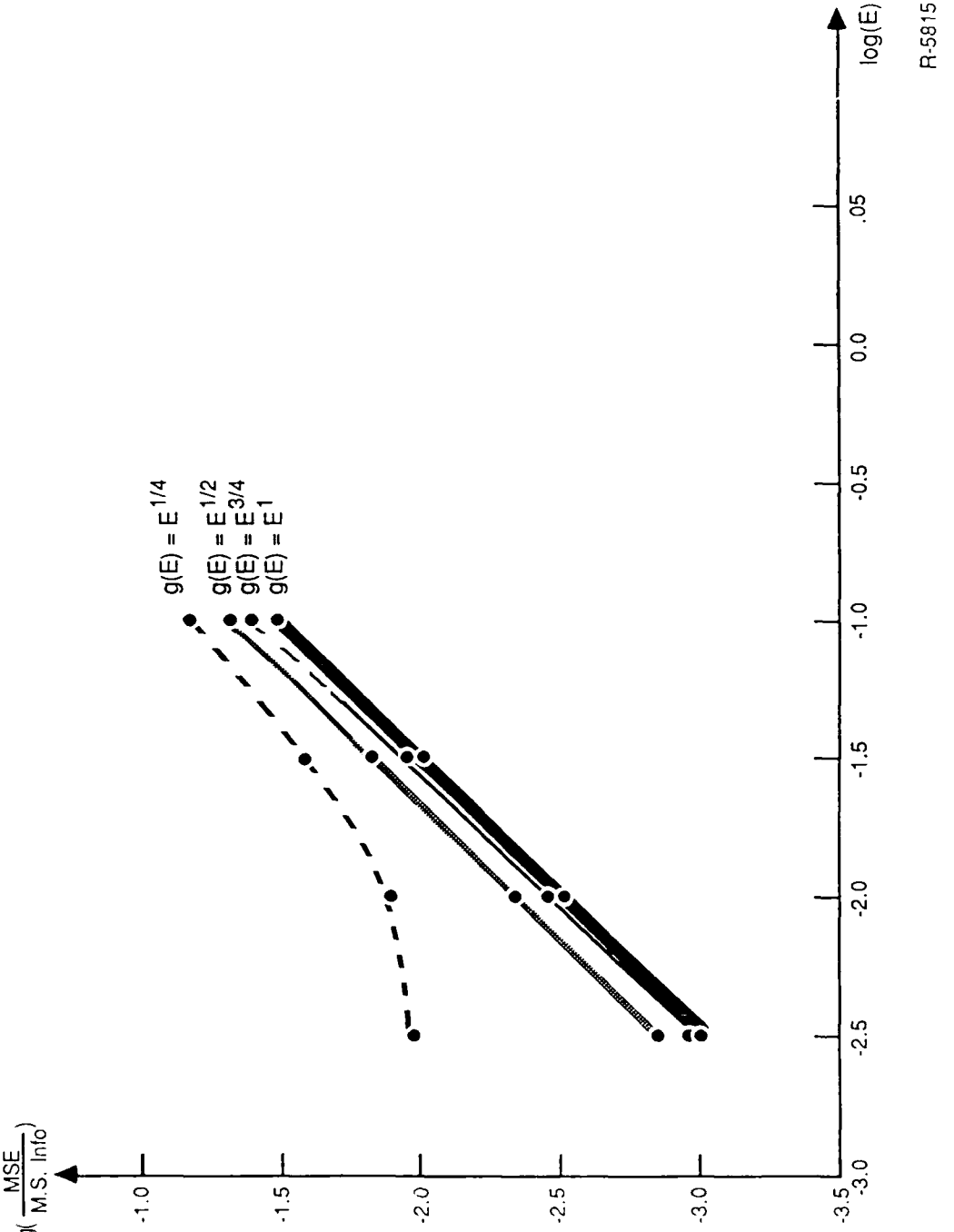


Figure 8-17. Plot of Log $\left(\frac{\text{Error}}{\text{Information}}\right)$ for Various Orders of $g(\epsilon)$

$$\text{Mean Square Information} = \frac{1}{2} \log \left(\frac{1}{T} \int_0^T \left(p_1^L - \frac{\gamma 2}{\gamma 1 + \gamma 2} \right)^2 dt \right) \qquad (8\text{-}39)$$

$$\text{Power of epsilon in signal} = \frac{\log(g(\varepsilon))}{\log(\varepsilon)} \quad . \qquad (8\text{-}40)$$

The plot shows a distinct break point in its behavior at $g(\varepsilon) = \varepsilon^{1/2}$. Specifically, for $g(\varepsilon) = \varepsilon^k$, $k < 1/2$, there is very little difference in the mean square information rate generated by changing the value of ε over several orders of magnitude. For values of $k > 1/2$, there is a clear drop in the information available to the filter as ε decreases. This supports our analysis which indicated that $g(\varepsilon) = \varepsilon^{1/2}$ is a critical dividing point in the performance of these estimation algorithms.

The second graph plots the logarithm of the mean square error against the power of ε in $g(\varepsilon)$. In this case we do not see the threshold effect and the plots appear linear for the values of ε that were simulated. When the two sets of data are combined and the quotient plotted against ε, we see a change in behavior at $g(\varepsilon) = \varepsilon^{1/2}$. For $g(\varepsilon) < \varepsilon^{1/2}$, the plots are linear, which slope 1, indicating a linear relationship with ε as predicted in Section 7. For $g(\varepsilon) > \varepsilon^{1/2}$, the error is much larger relative to the available information and there is no evidence that the ratio of error to information will decrease significantly more. This is again in agreement with the theoretical results of Section 7.

Supremum Results:

An identical set of graphs (Figs. 8-18 through 8-20) was prepared for estimates of the suprema of the information and the approximation error. The estimates were calculated using Eqs. 8-41 and 8-42 for the simulation data.

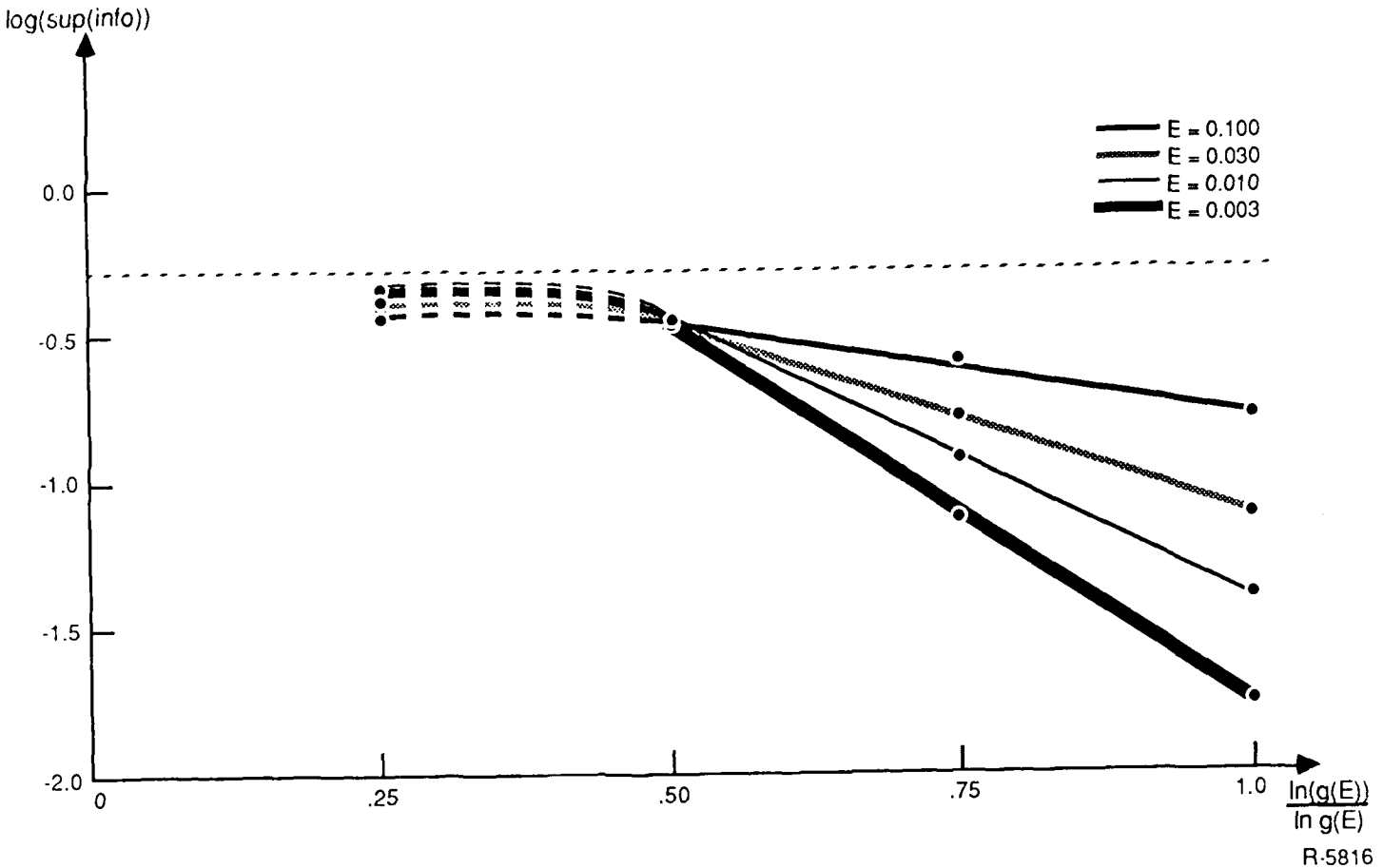


Figure 8-18. Plots of $\sup_t \left| p^L - \frac{\gamma_2}{\gamma_1+\gamma_2} \right|$ of Exact Filter

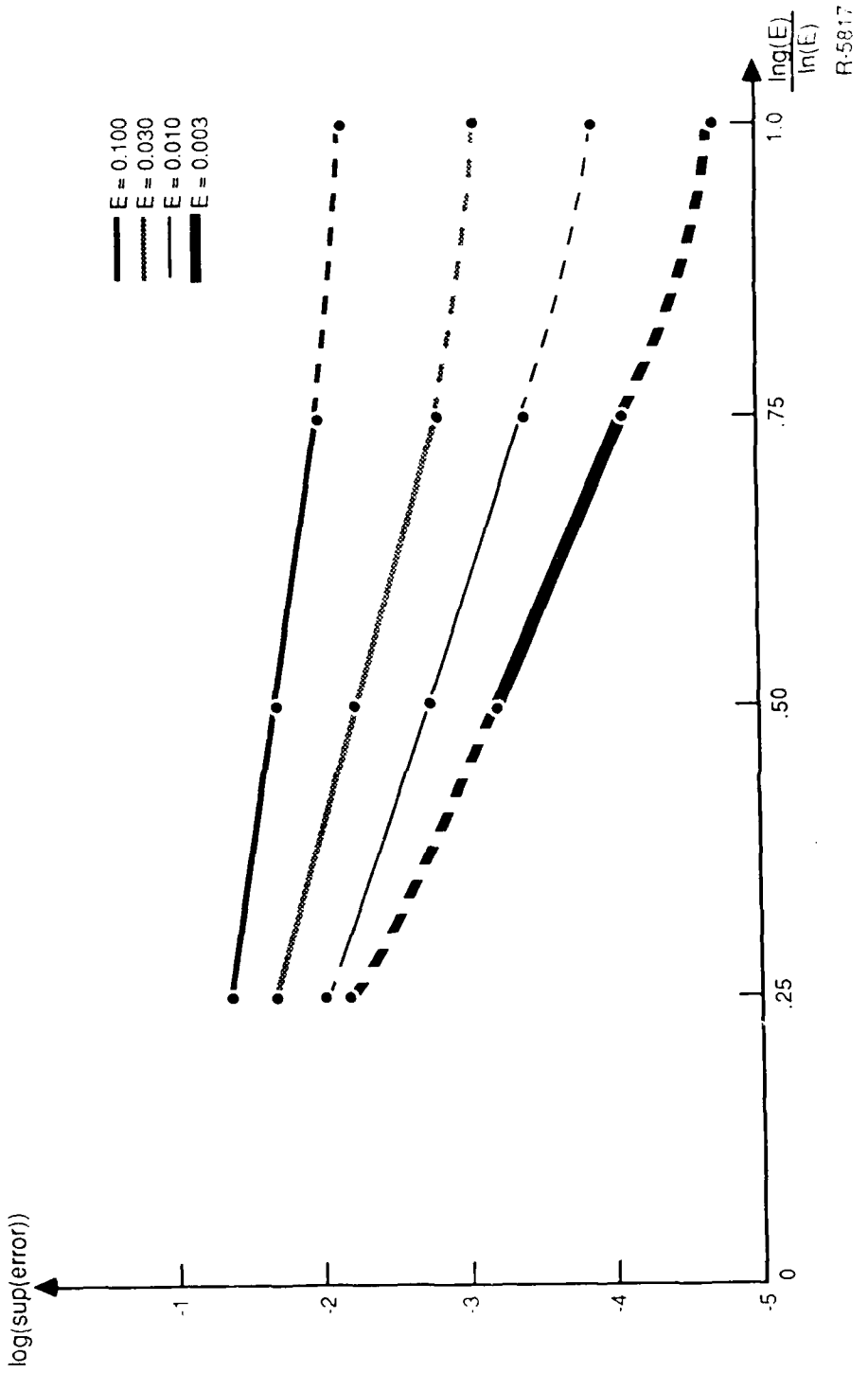


Figure 8-19. Plot of Log [sup(error)] for Approximate Filter in White Noise

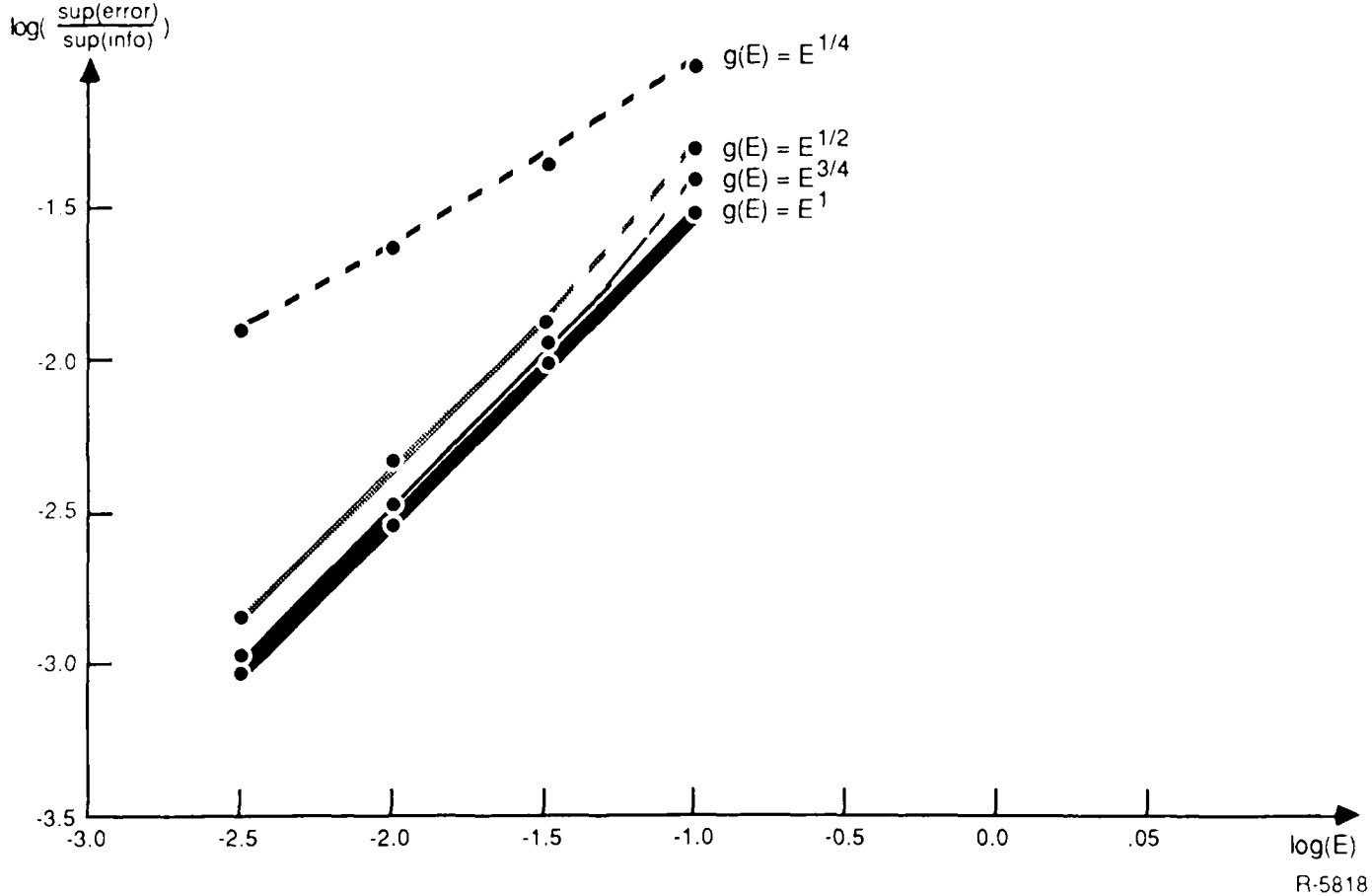


Figure 8-20. Plot of Log $\left(\frac{\sup\ |\text{error}|}{\sup\ |\text{info.}|}\right)$ for Various Orders of $g(\varepsilon)$

$$\hat{I}_{sup} = \max_t \left| p^L(t) - \frac{\gamma_2}{\gamma_1+\gamma_2} \right| \tag{8-41}$$

$$\hat{E}_{sup} = \max_t \left| p^L(t) - p_L(t) \right| \tag{8-42}$$

where $p^L(t)$ and $p_L(t)$ are defined as the exact and approximate probabilities respectively. The first graph provides a plot of log $(\hat{I})$ versus the power of ϵ in $g(\epsilon)$. The plot exhibits identical characteristics to that for the mean square quantity. Comparing plots of maximum error and the quotient of error and information, we find that they are also very similar. The last plot, the quotient of the error and the information, exhibits a slope of 1 as in the mean square case.

8.3.3 Sample Paths and Probabilities for the Discrete Measurement Case

The discrete measurement simulations were run with three different filters as described in subsection 8.2. The set of nominal parameter values that were chosen for the simulation are given in Eq. 4-43.

$$\epsilon = 0.01 \tag{8-43a}$$

$$\lambda_1 = 1 + g(\epsilon) \tag{8-43b}$$

$$\lambda_2 = \lambda_3 = \lambda_4 = 1 \tag{8-43c}$$

$$\lambda_1 = \lambda_2 = \mu_3 = \mu_4 = 1 \tag{8-43d}$$

$$\Delta t = 0.1 \tag{8-43e}$$

I. EXACT FILTER

Simulations of the exact filter were run for $g(\varepsilon)$ ranging from 0.03 to 3.0. Prior to the simulations we expected very poor performance for $g_1(\varepsilon) < \varepsilon^{1/2}$, marginal performance for $g(\varepsilon) = \varepsilon^{1/2}$ and excellent performance for $g(\varepsilon) > \varepsilon^{1/2}$. The simulations are in fairly close agreement with these conjectures. In the first plot (Fig. 8-21) for $g_1(\varepsilon) = 0.03$, the filter output appears much like a random walk about the $p_L = 0.5$ line. As $g_1(\varepsilon)$ is increased to 0.1, or $\varepsilon^{1/2}$, in Fig. 8-22, performance improves with the filter reaching the correct conclusion regarding the correct state of the system when the system remains in that state for times which are $O(\varepsilon^{-1})$. The filter has little chance, however, of detecting transitions to the left or right if the duration of such a sojourn is for a significantly shorter period of time. As the value of $g_1(\varepsilon)$ is increased to .3, 1, and finally 3 (Figs. 8-23 through 8-25), the filter output becomes much better, and begins displaying "switching behavior." In addition, short excursions of the system to the left and right are detected.

II. DIFFERENTIAL EQUATION APPROXIMATION TO DISCRETE MEASUREMENT FILTER

The differential equation approximations to the optimal filter was simulated with identical nominal parameter values as in I, but with $g(\varepsilon) = 0.1$ and time intervals for integration of 1, 10 and 50 seconds. In our integration technique for this case (driven by a jump process), second order effects which are $O(g^2(\varepsilon)\ \Delta K^2)$ were ignored. Since $\lambda_i \cong \lambda_i^{-1} \cong 1$ for all of our simulations $O(\Delta K) = O(T(\varepsilon))$ and therefore ignoring the second order terms is justified for $g(\varepsilon)\ T(\varepsilon)$ small. In our simulations, $g(\varepsilon)$ was 0.1, so for the case of $T(\varepsilon) = 1$ (Fig. 8-26), the differential equation should provide a good approximation, for $T(\varepsilon) = 10$ (Fig. 8-27) a marginal approximation, and for $T(\varepsilon) = 50$ (Fig. 8-28)

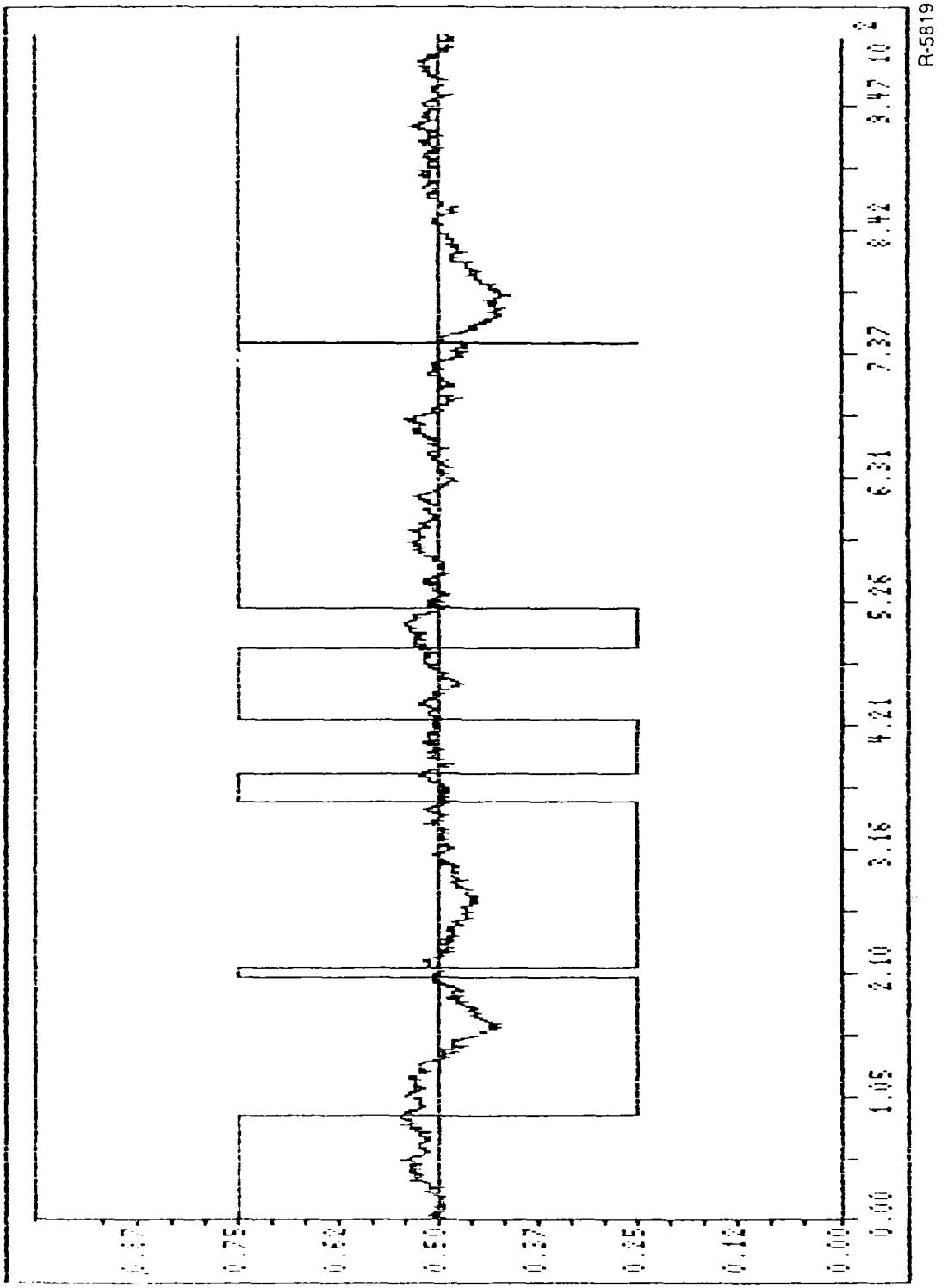


Figure 8-21. Exact Probabilities for Discrete Model, $g_1(\varepsilon) = 0.03$, $\varepsilon = 0.01$, $\Delta t = 0.1$

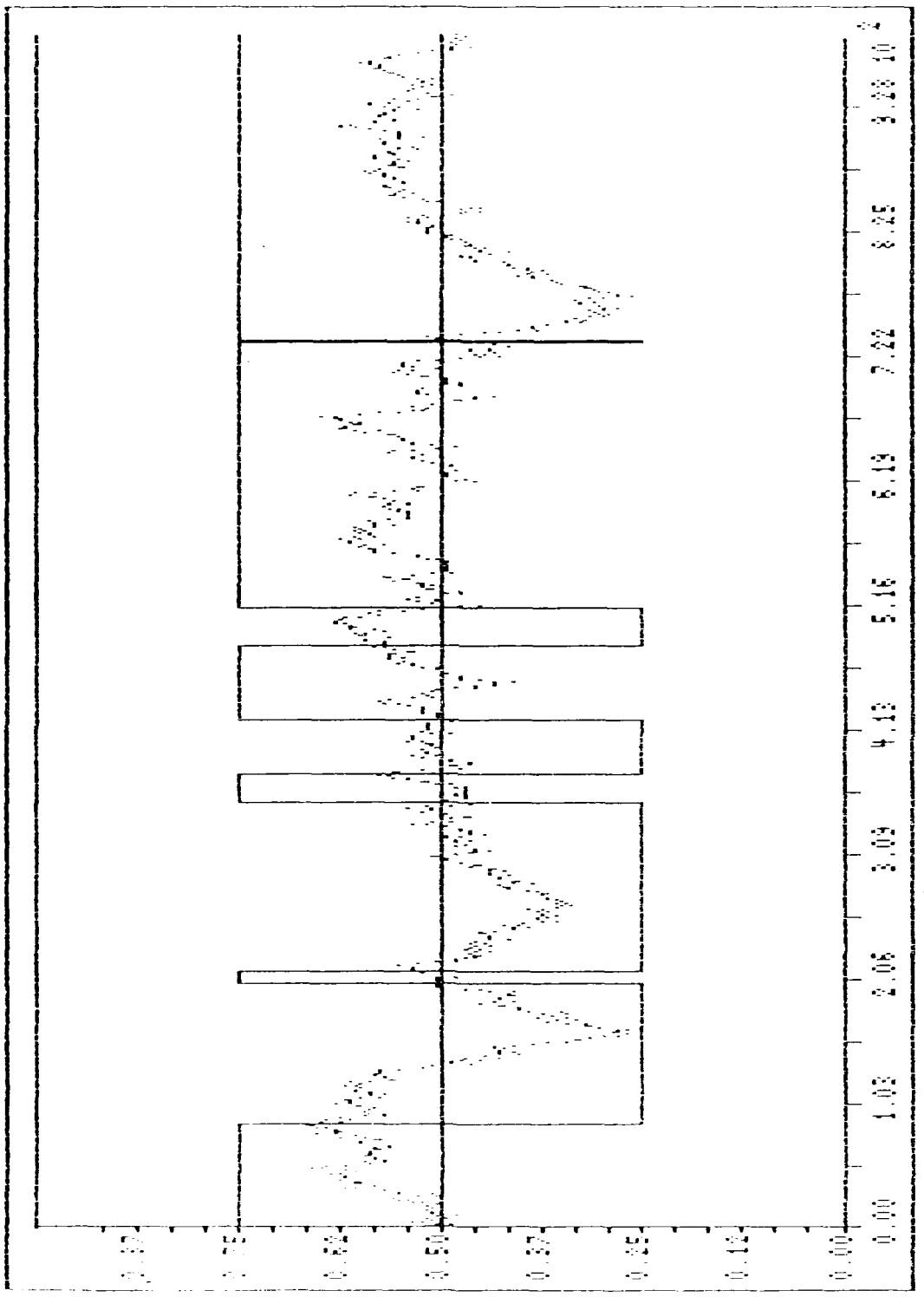


Figure 8-22. Exact Probabilities for Discrete Model, $g_1(\epsilon) = 0.1$, $\epsilon = 0.01$, $\Delta t = 0.1$

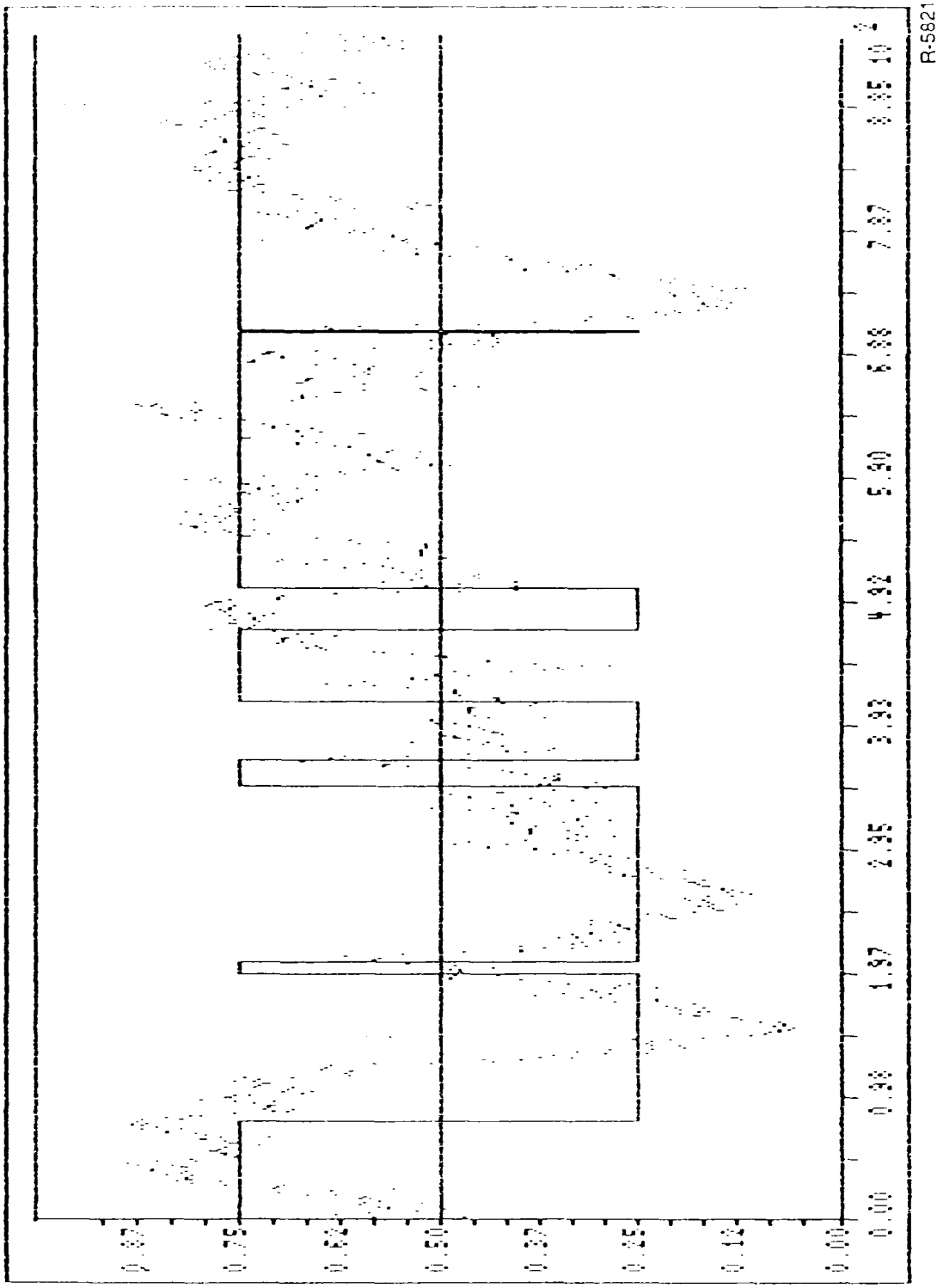


Figure 8-23. Exact Probabilities for Discrete Model, $g_1(\epsilon) = 0.3$, $\epsilon = 0.01$, $\Delta t = 0.1$

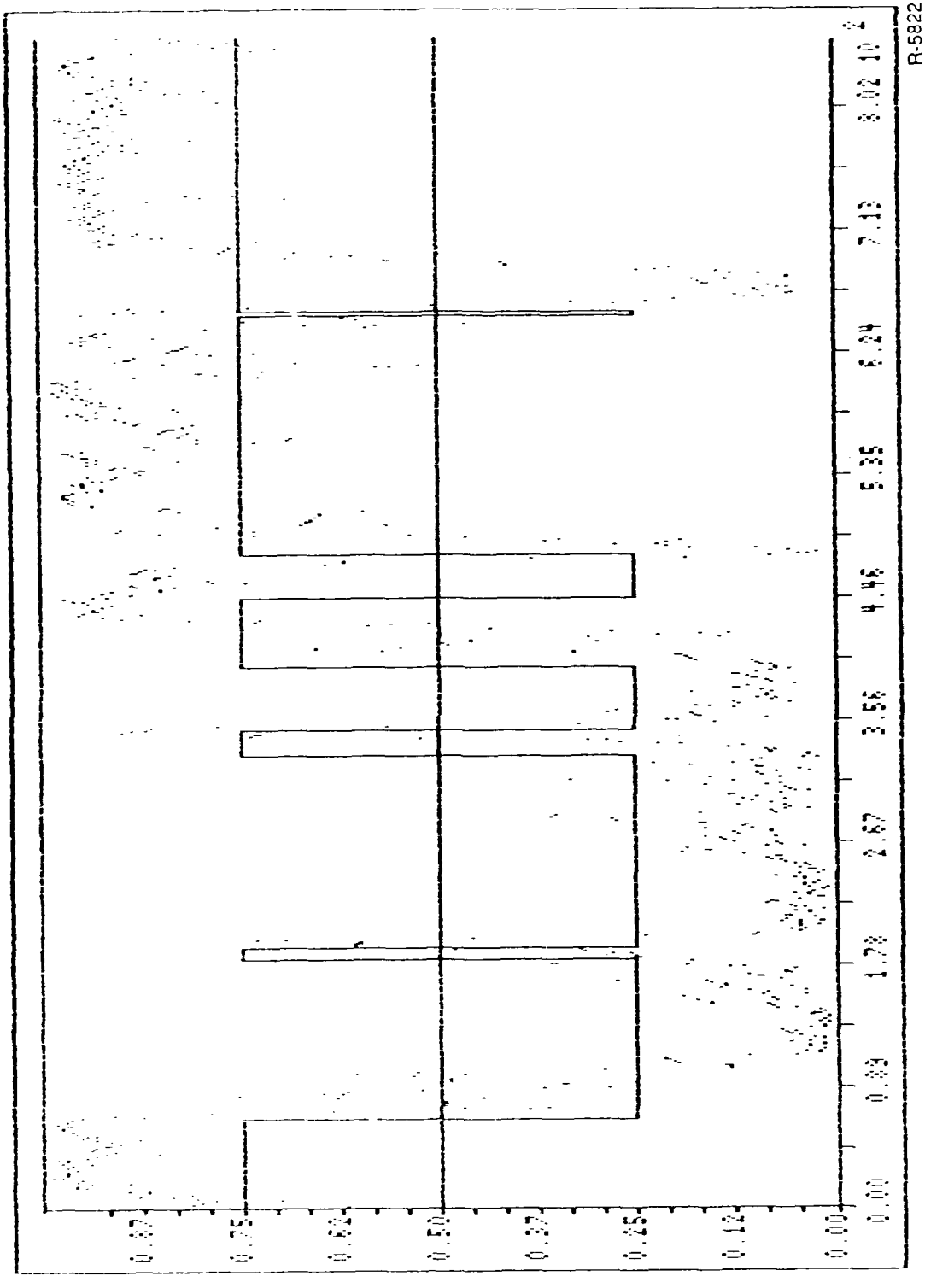


Figure 8-24. Exact Probabilities for Discrete Model, $g_1(\epsilon) = 1.0$, $\epsilon = 0.01$, $\Delta t = 0.1$

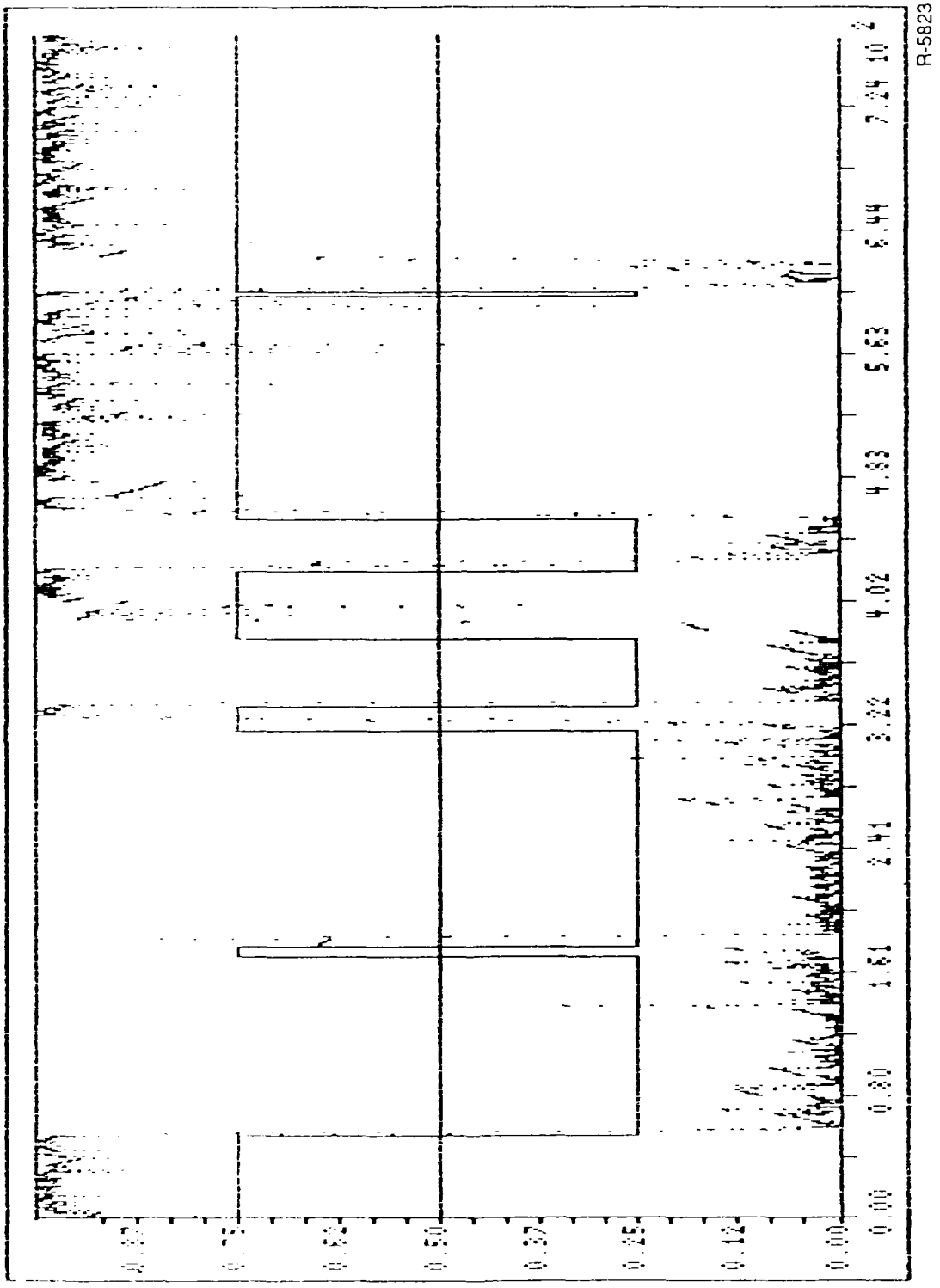


Figure 8-25. Exact Probabilities for Discrete Model, $g1(\epsilon) = 3.0$, $\epsilon = .01$, $\Delta t = 0.1$

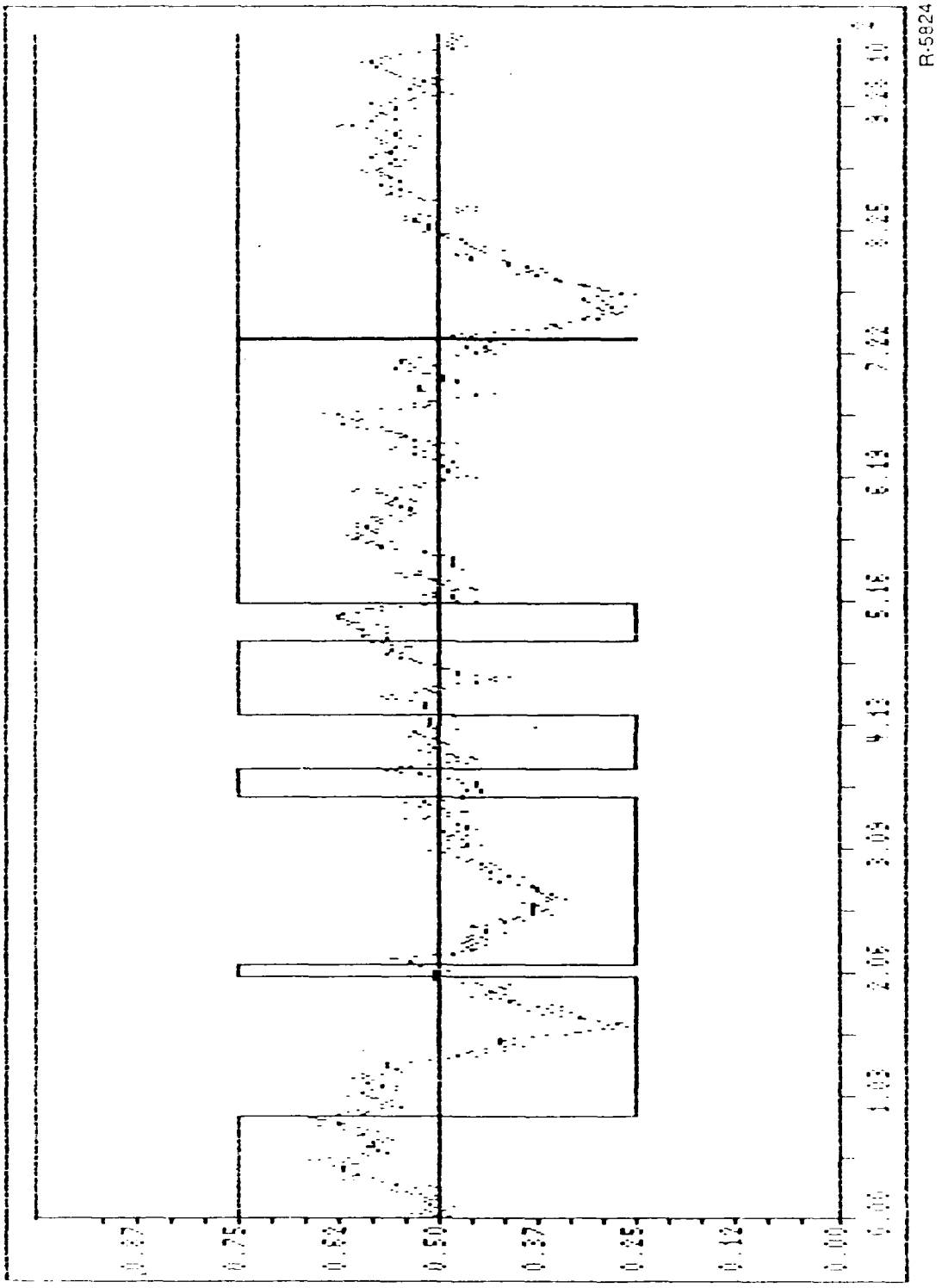


Figure 8-26. Differential Equation Approximation with Time Increment of 1

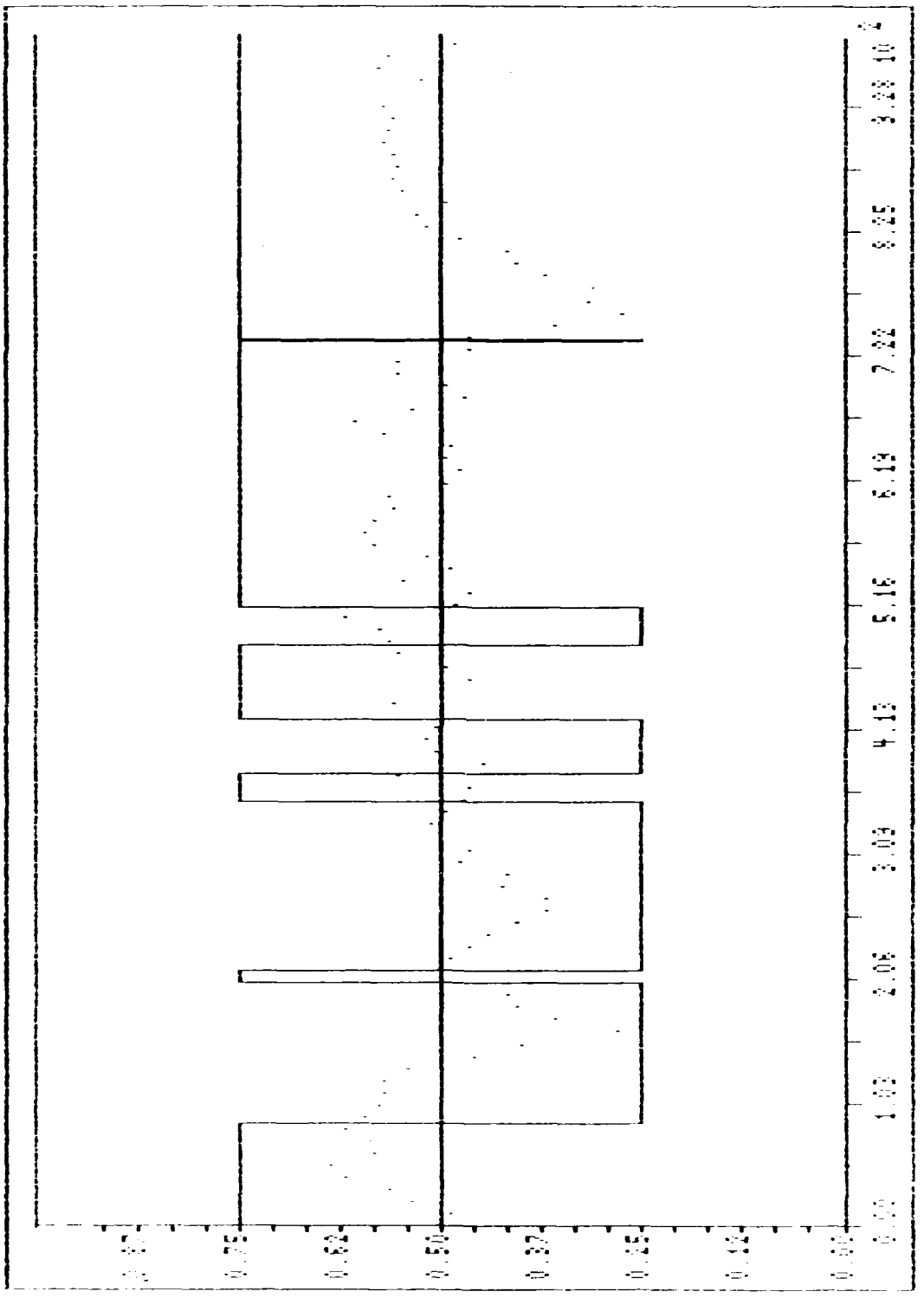


Figure 8-27. Differential Equation Approximation with Time Increment of 10

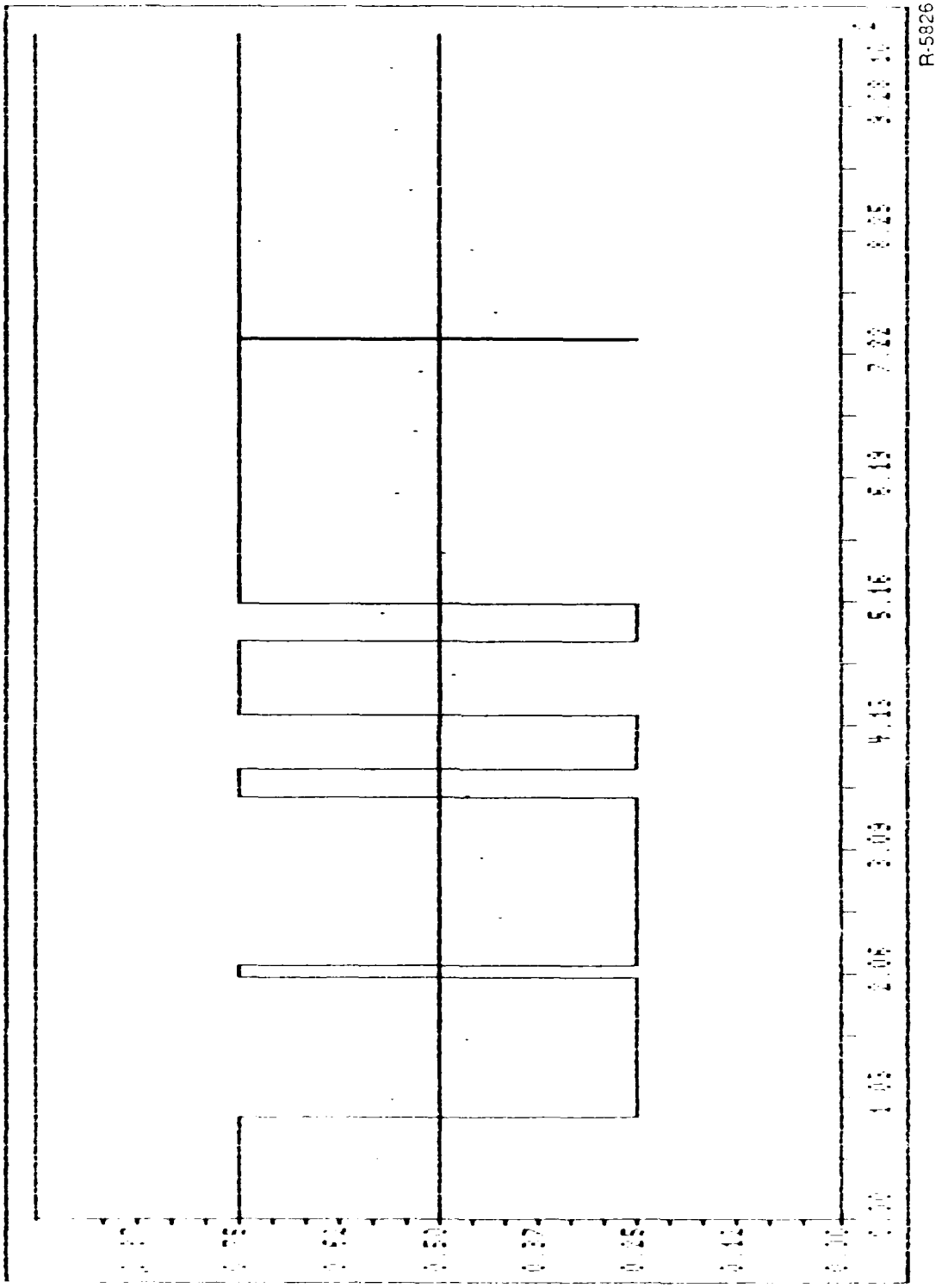


Figure 8-28. Differential Equation Approximation with Time Increment of 50

a poor approximation. Examining the simulation results in these figures and comparing them to Fig. 8-22, we see that the maximum deviations from optimum are approximately 0.01, 0.04 and 0.12, respectively, confirming our expectations. To provide further support, two additional simulation results are provided both with $T(\varepsilon) = 10$, but in the first case, $g(\varepsilon) = 0.03$ for which we expect good performance, and in the second, $g_1(\varepsilon) = 0.3$, for which we expect poor performance (compared to optimal). The plots of the simulations are provided (Figs. 8-29 and 8-30) and support these expectations, with maximum deviations of 0.005 and 0.150 for these sample paths.

III. FE/BE STRUCTURE FOR DISCRETE MEASUREMENT MODEL

The FE/BE filter structure was simulated with different lengths of time used for the batching interval at the front end. For this filter we expect good performance relative to optimum for cases with $g(\varepsilon)\, T(\varepsilon) < 1$ and poorer performance when $g(\varepsilon)\, T(\varepsilon) > 1$. (Note: This is only a rough threshold, a more precise condition on the parameters is conjectured in subsection 8.3.4. Therefore, since a value of $G(\varepsilon) = 0.01$ was used in this section, we expect reasonable performance for $T(\varepsilon)$ being 50 or even higher. This was the case in the simulations with maximum deviations of .01, .02, .04 from optimal for $T(\varepsilon) = 1$, 10 and 31, respectively (Figs. 8-31 through 8-33). In the final simulation for $g(\varepsilon) = 0.1$ (Fig. 8-34), $T(\varepsilon)$ was set to 75 and therefore we expect that we will see marginal performance ($\varepsilon T(\varepsilon) = 0.75$). The simulations showed deviations by as much as .08 from optimal but in general agreement was quite good. A final simulation with $g(\varepsilon) = 0.3$ and $T(\varepsilon) = 20$ was run to demonstrate the case where absolute filter performance is fairly good. In this case $\varepsilon T(\varepsilon)$ was 0.2 and relative performance was very good as well (compare Fig. 8-35 to Fig. 8-23).

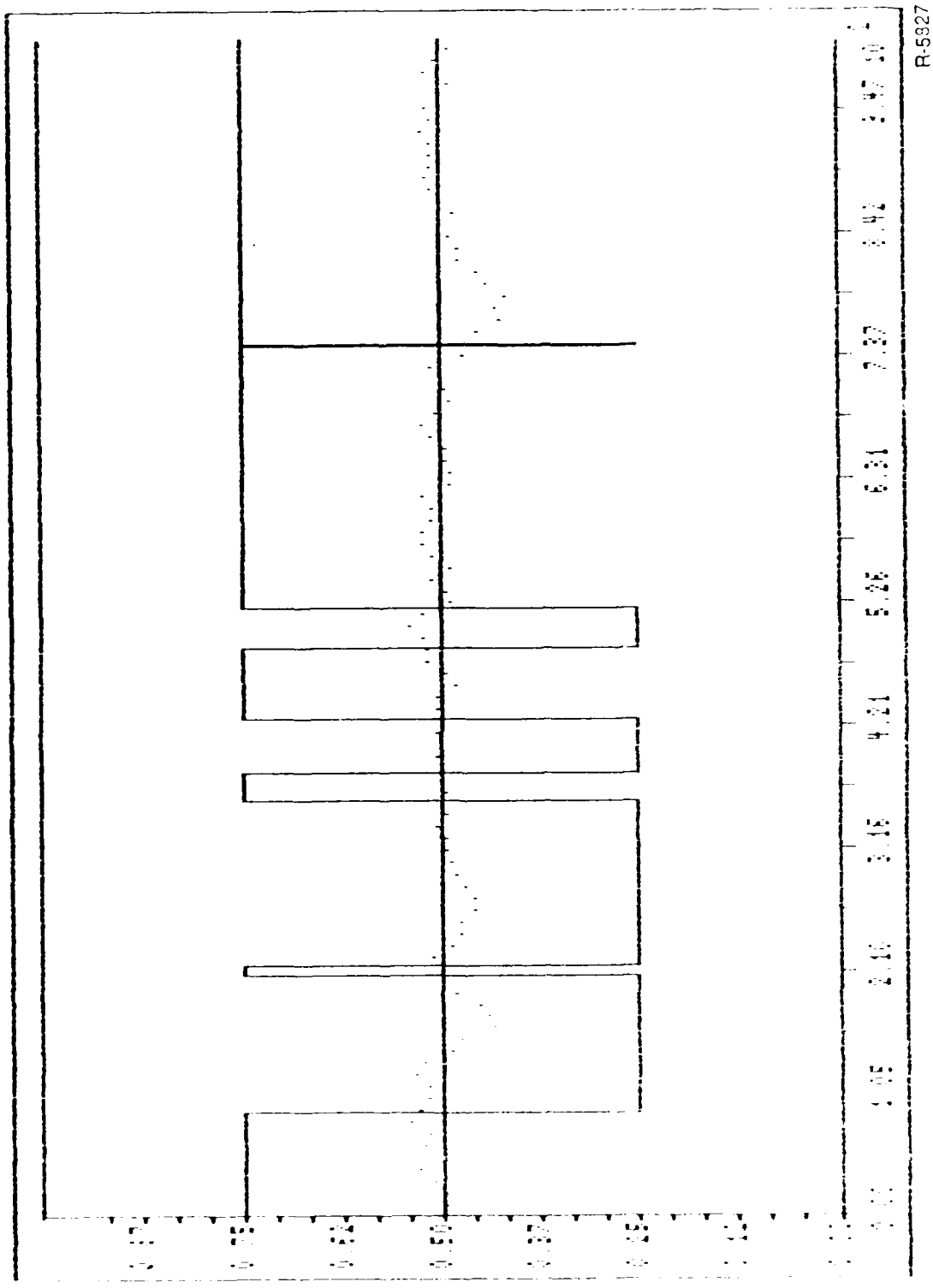


Figure 8-29. Differential Equation Approximation with Time Increment of 10 and $g_1(\epsilon) = 0.03$

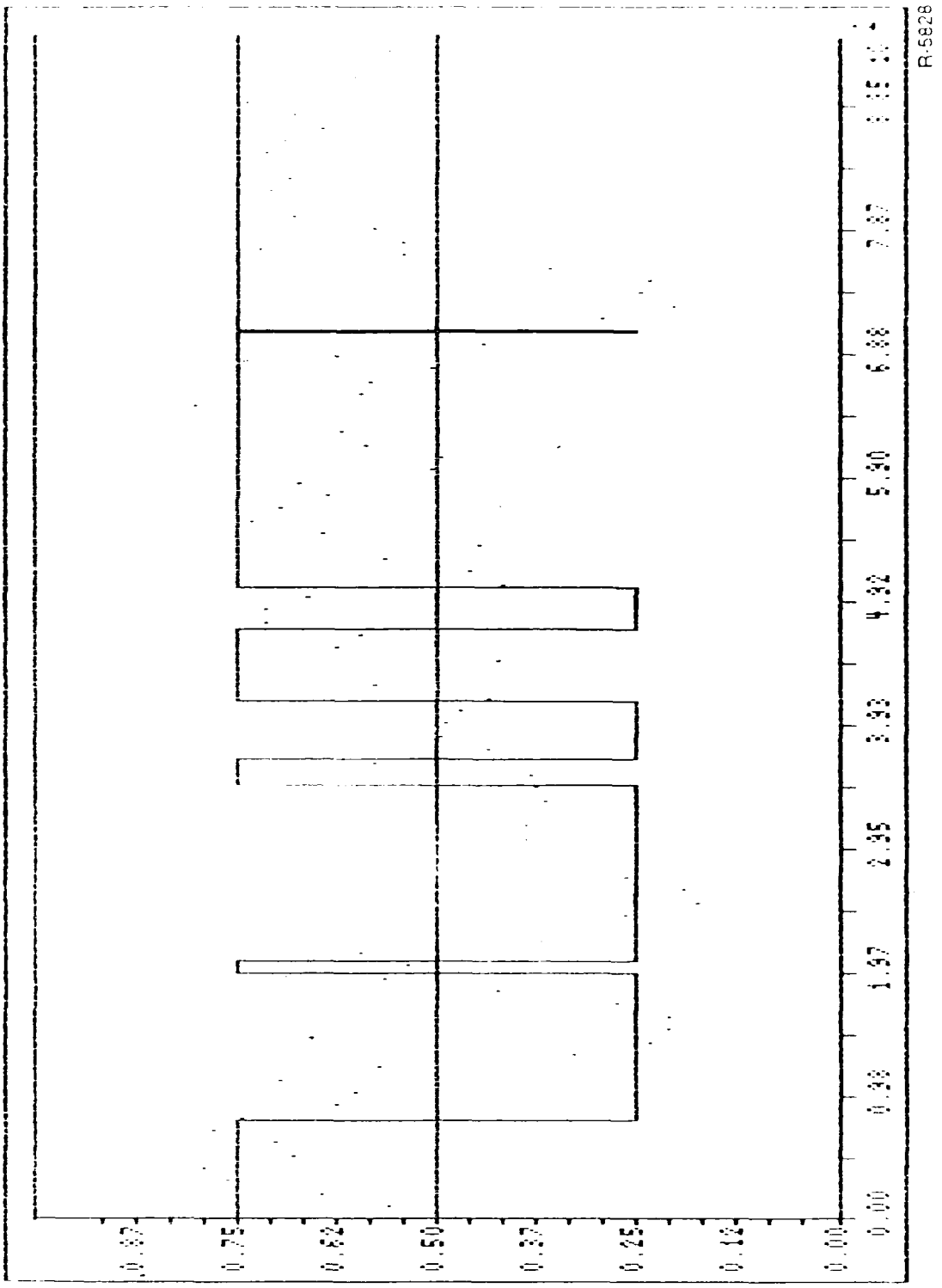


Figure 8-30. Differential Equation Approximation with Time Increment of 10 and $g_1(\epsilon) = 0.3$

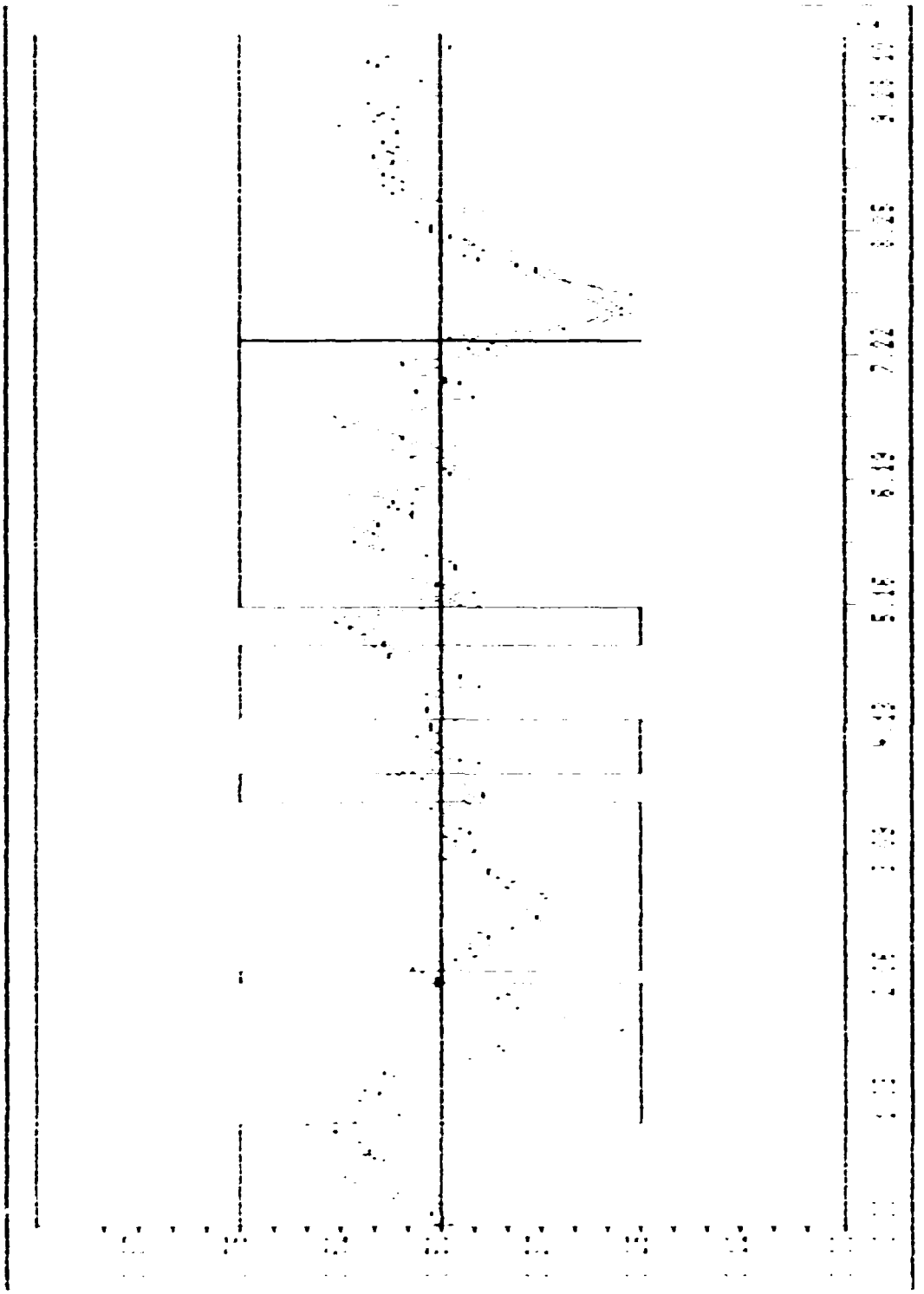


Figure 8-31. FE/BE Simulation for $g_1(\epsilon)$ = 0.1 and Updates Every 1 Second

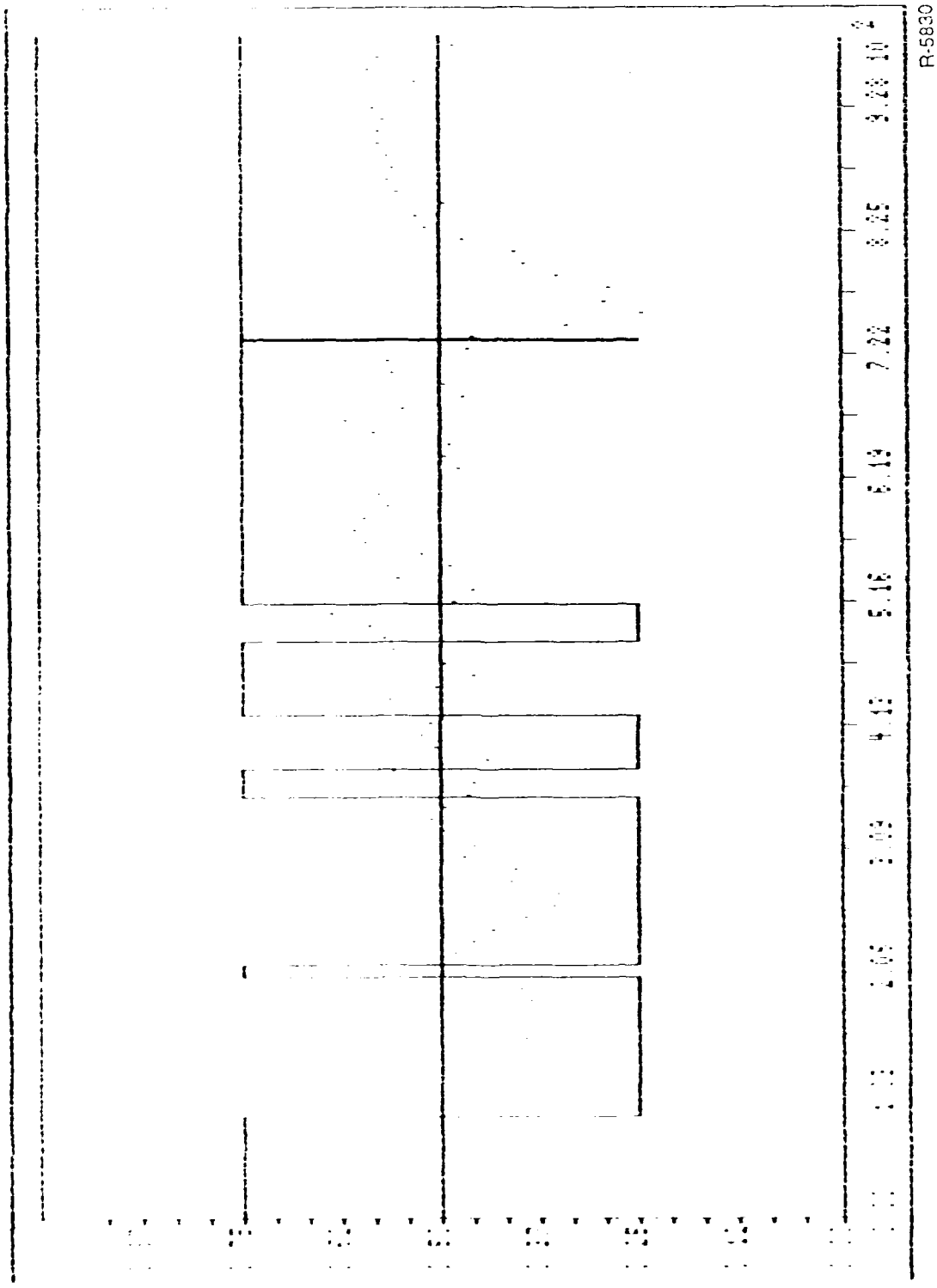


Figure 8-32. FE/BE Simulation for $g_1(\epsilon)$ = 0.1 and Updates Every 10 Seconds

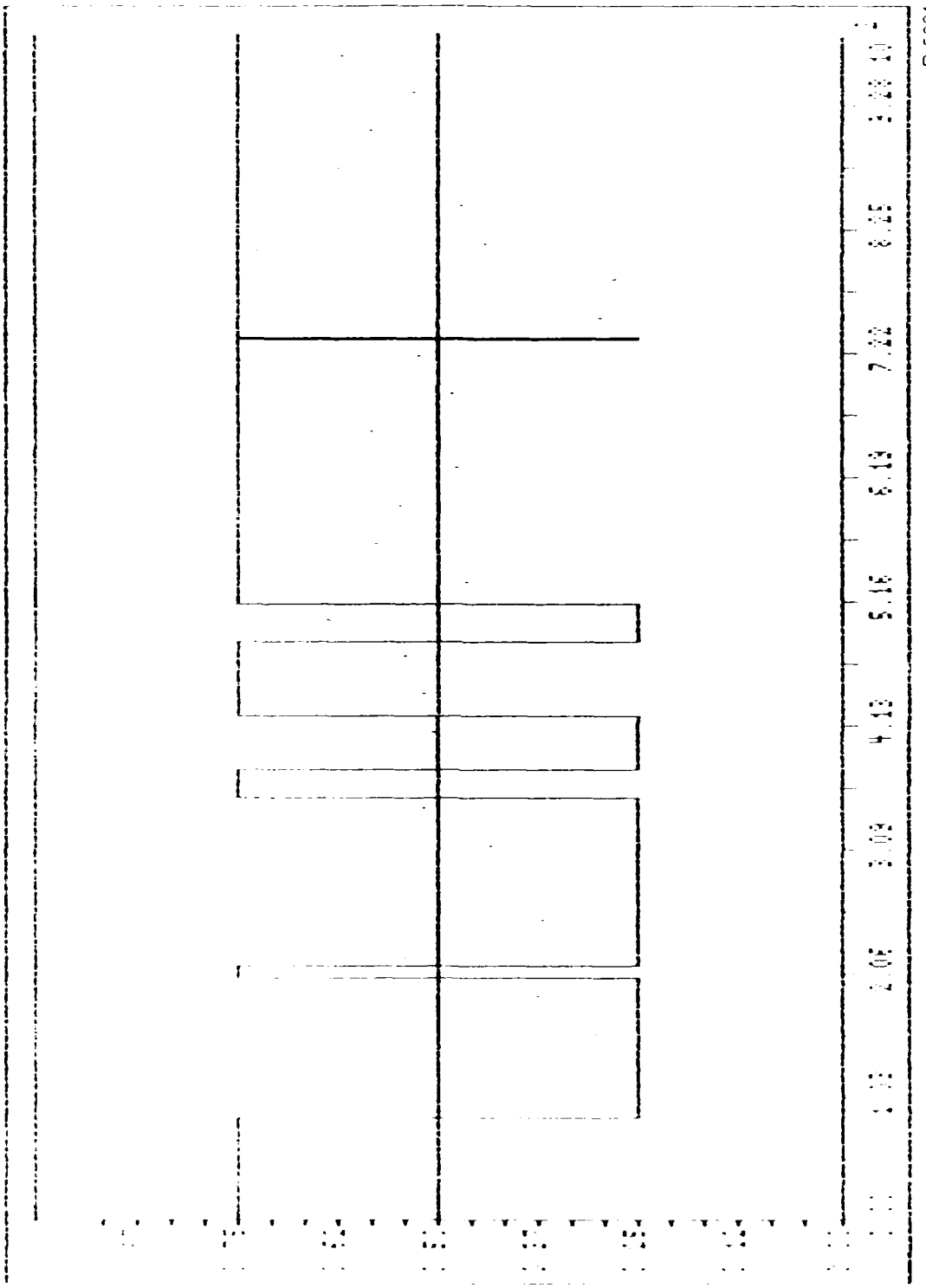


Figure 8-33. FE/BE Simulation for $g_1(\epsilon)$ = 0.1 and Updates Every 31 Seconds

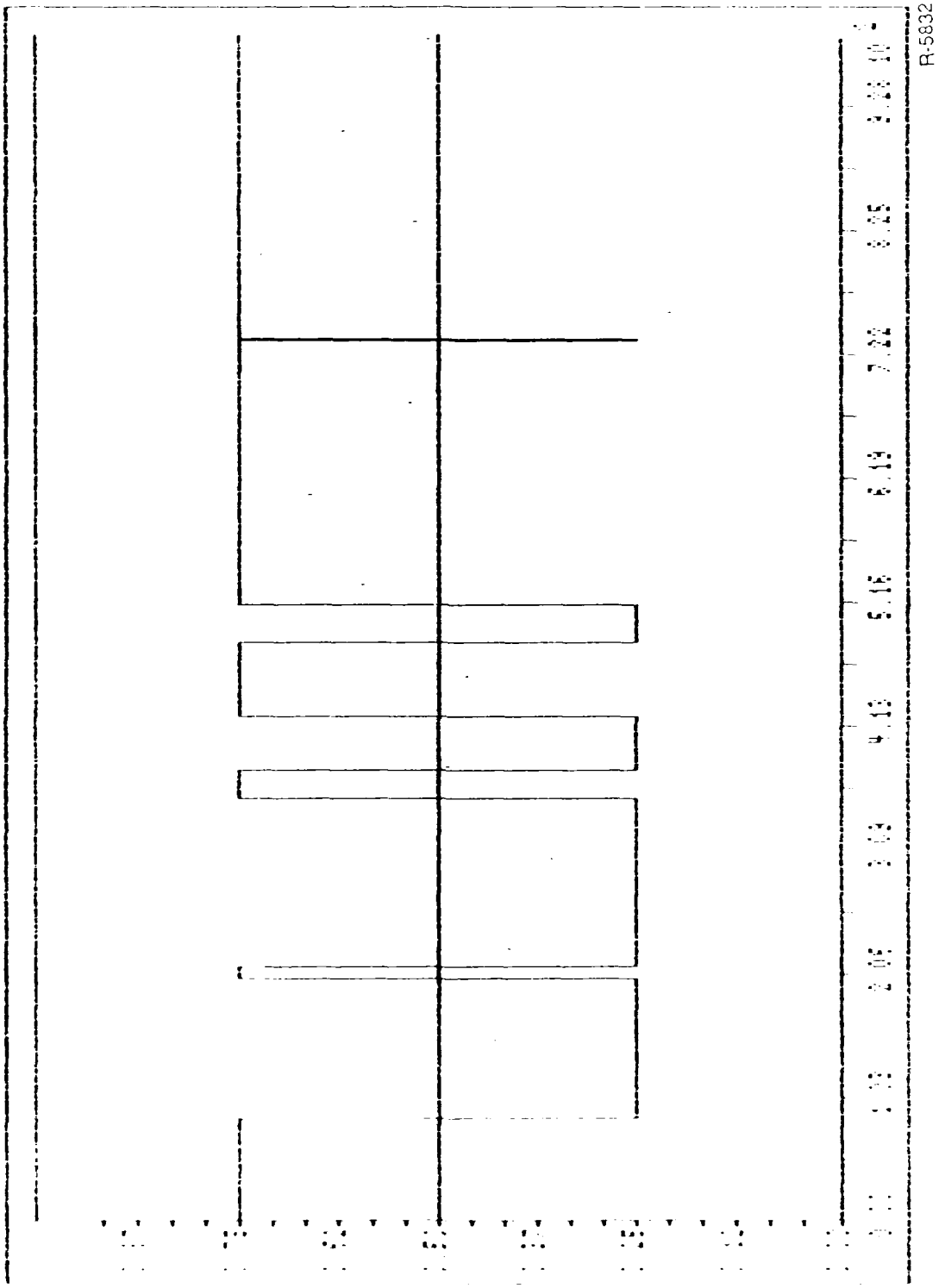


Figure 8-34. FE/BE Simulation for $g_1(\epsilon)$ = 0.1 and Updates Every 75 Seconds

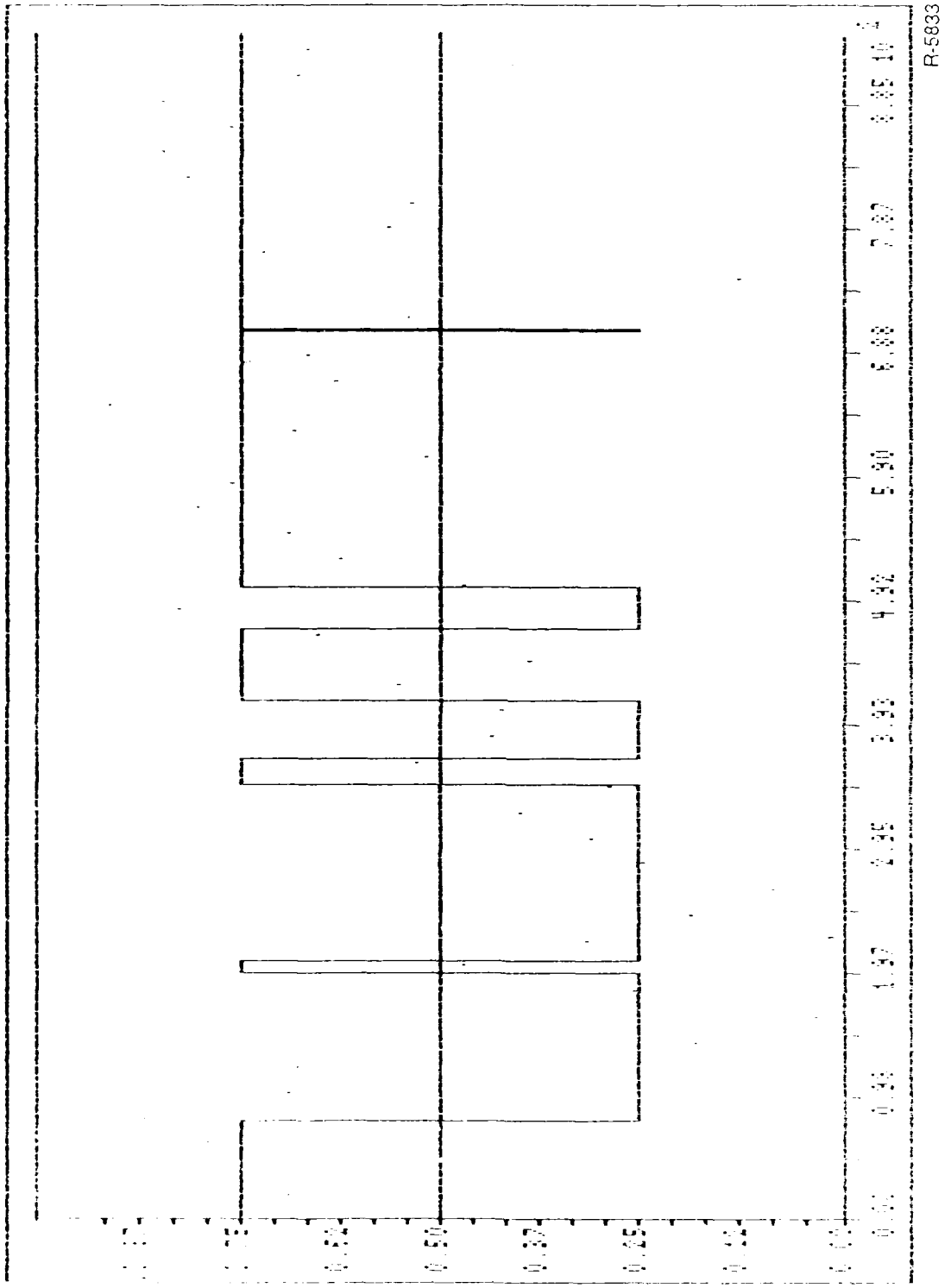


Figure 8-35. FE/BE Simulation for $g_1(\epsilon) = 0.3$ and Updates Every 20 Seconds

8.3.4 Quantitative Simulations for FE/BE Processor in a Noiseless Environment

In a similar manner to subsection 8.3.2, a number of runs were performed for the filter which approximated the optimal filter in the discrete measurement case by calculating a statistic L(t) in a batch fasion and using this in a slower set of calculations. The results are provided in Figs. 8-36 to 8-45. The first two graphs (Figs. 8-36 and 8-37) provide log-log plots of the difference between optimal and approximate filters versus $g(\varepsilon)$ for the value of $T(\varepsilon)$, the time between batches, varying from 1 to 31 units. The value of ε is selected to be 0.1 in the first case and 0.001 in the second. These plots provide evidence that:

1. The difference between approximation and optimal is linear in $g^2(\varepsilon)$.
2. For $T(\varepsilon)$ small relative to ε^{-1}, changes in $T(\varepsilon)$ should have little effect on relative filter performance.

Figure 8-38 and 8-39 confirm the second point by displaying approximation error versus $T(\varepsilon)$ for ε ranging from 0.001 to 0.1. Finally, using these plots, the mean square approximation error was conjectured to be of the following form:

$$\lim_{\substack{T\to\infty \\ \varepsilon\to 0}} \frac{1}{T} \int_0^T \left(p^L(t) - p_L(t)\right)^2 dt = K_1 g^2(\varepsilon) \left(1 + K_2\, \varepsilon T^2(\varepsilon)\right) \qquad (8\text{-}44)$$

where K_1 and K_2 are constants. To support the conjecture, log-log plots of approximation error versus the expression in Eqs. 8-44 with $K_1 = K_2 = 1$ were generated for the simulation runs with ε = .1, .01, .001 (Figs. 8-40 through 8-42). The linearity of the plots and their slope of 1 support the conjectured formula in Fig. 8-42. In the case of ε = 0.001, the plots are somewhat

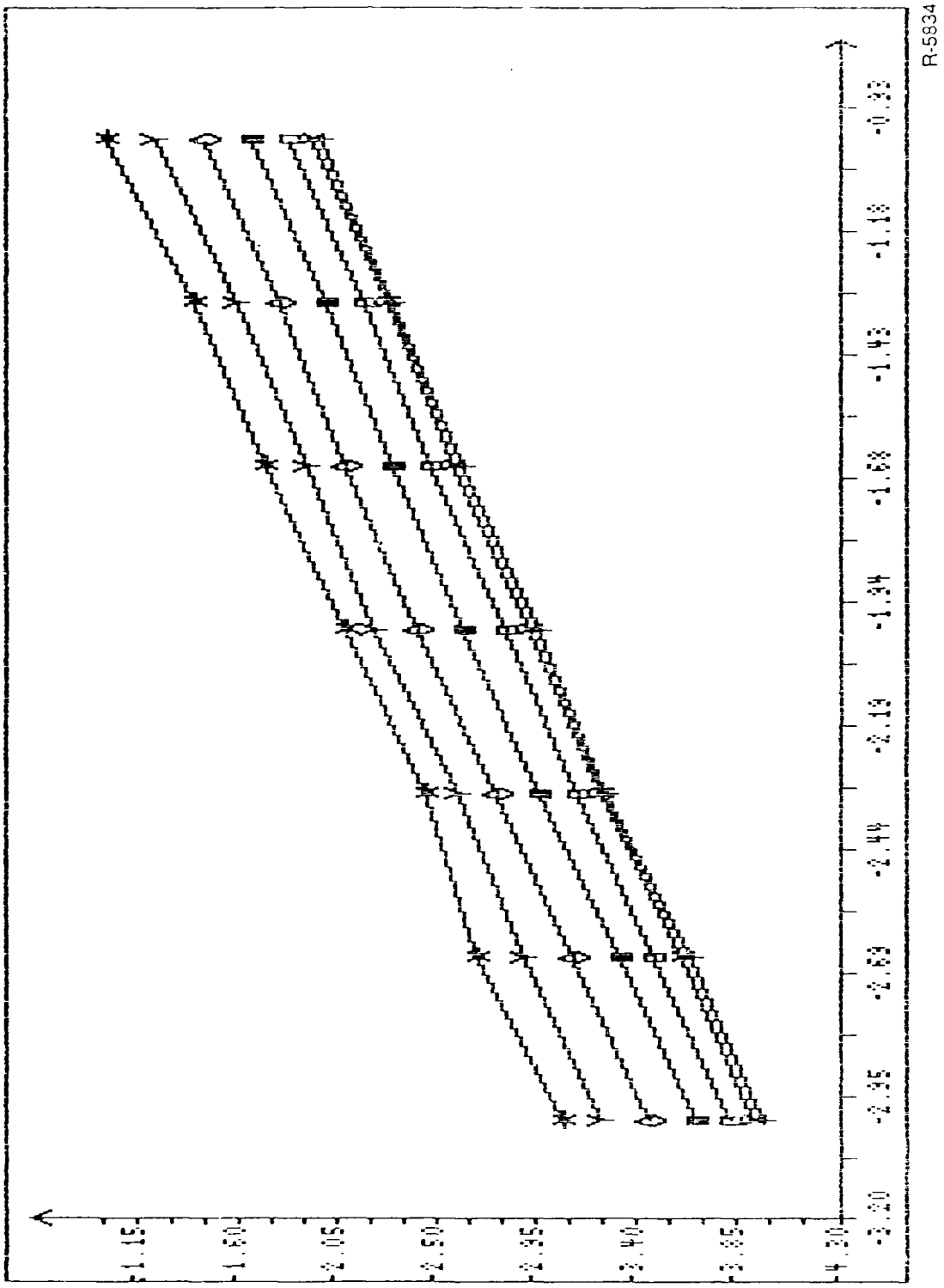


Figure 8-36. Plot of MSE vs. $g(\varepsilon)$ for $T(\varepsilon) = 1$ to 31 and $\varepsilon = 0.1$

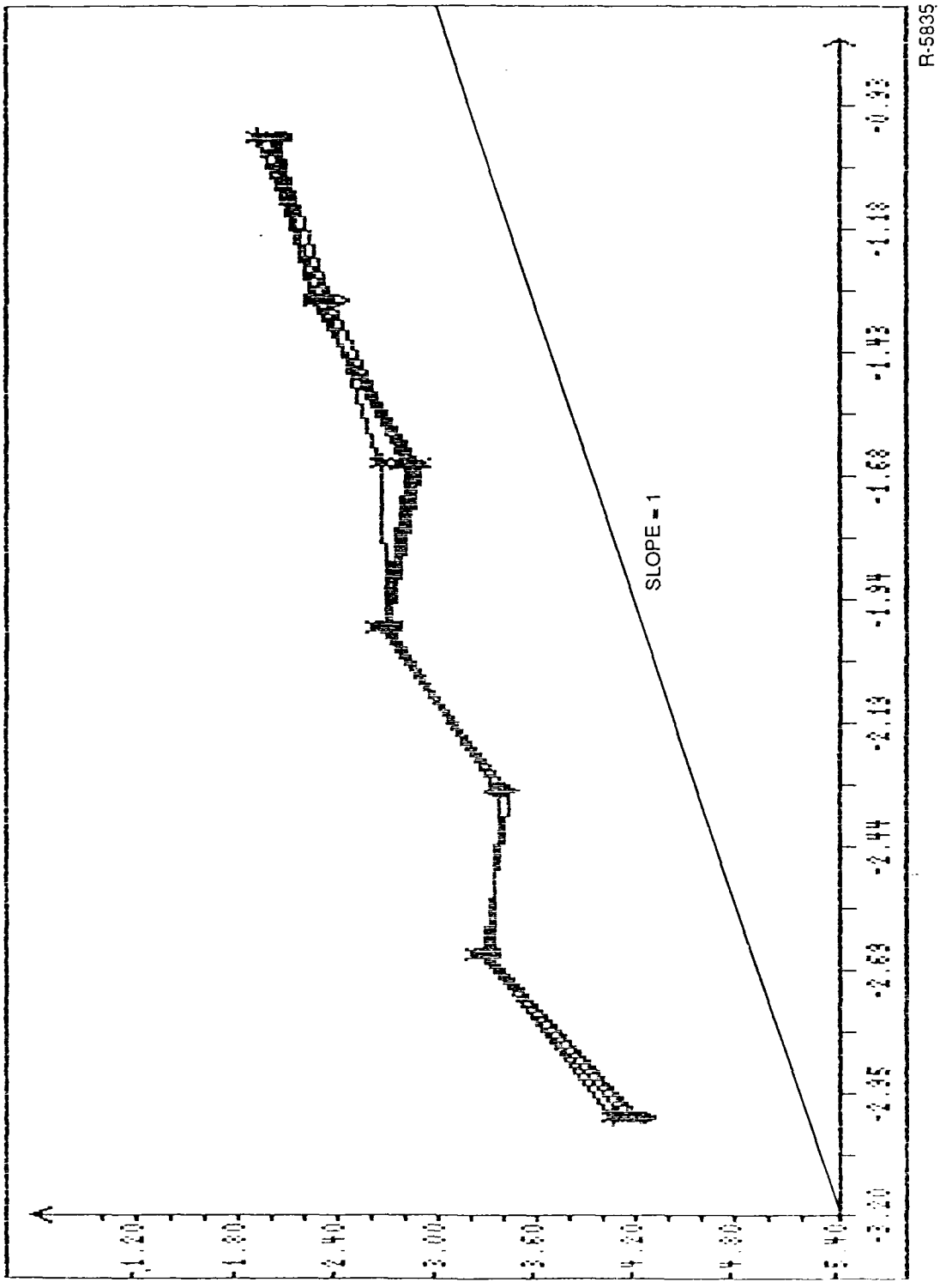


Figure 8-37. Plot of MSE vs. g(ε) for T(ε) = 1 to 31 and ε = .001

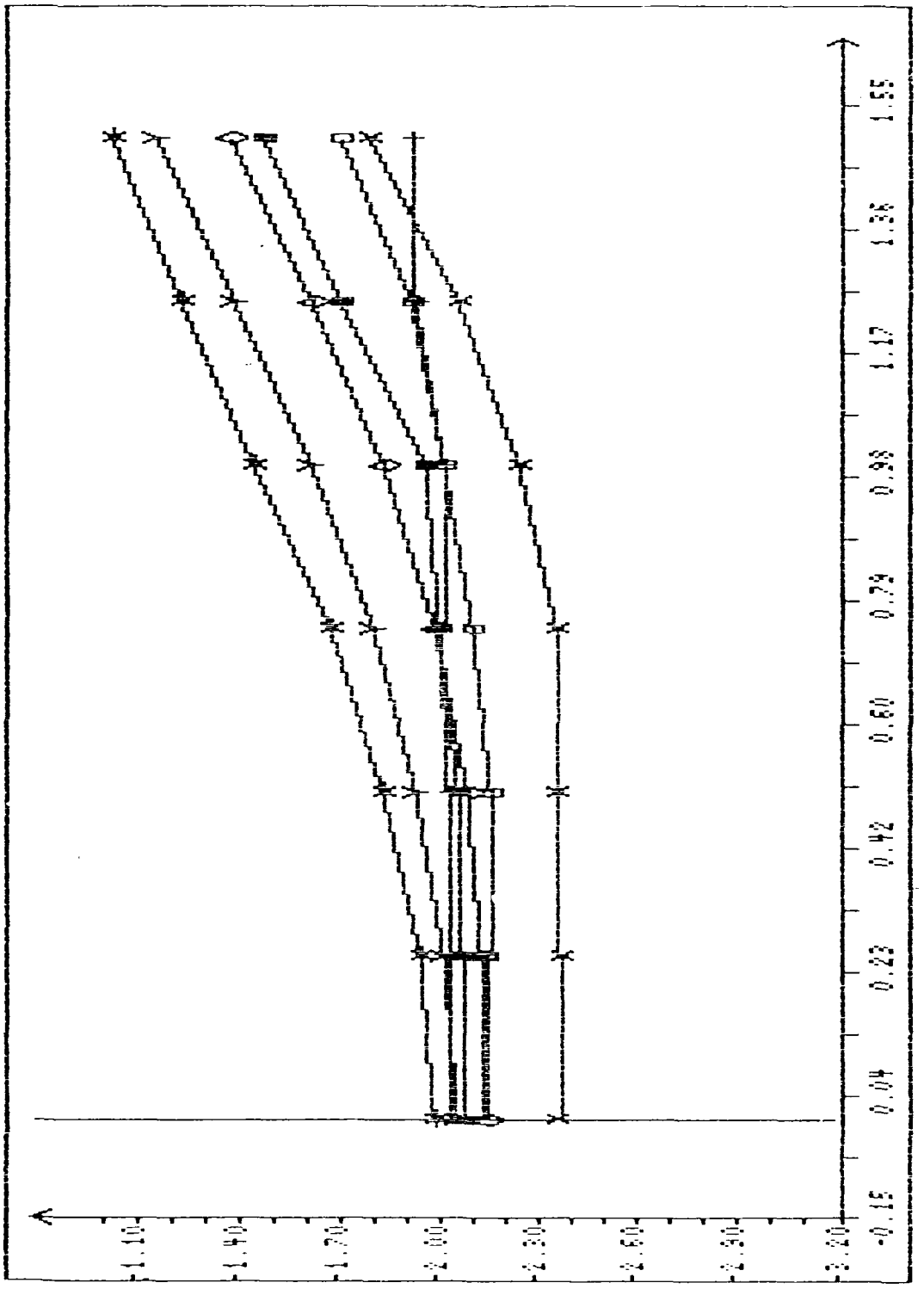


Figure 8-38. Plot of MSE vs. $T(\varepsilon)$ for $\varepsilon = .001$ to .1 and $g(\varepsilon) = 0.1$

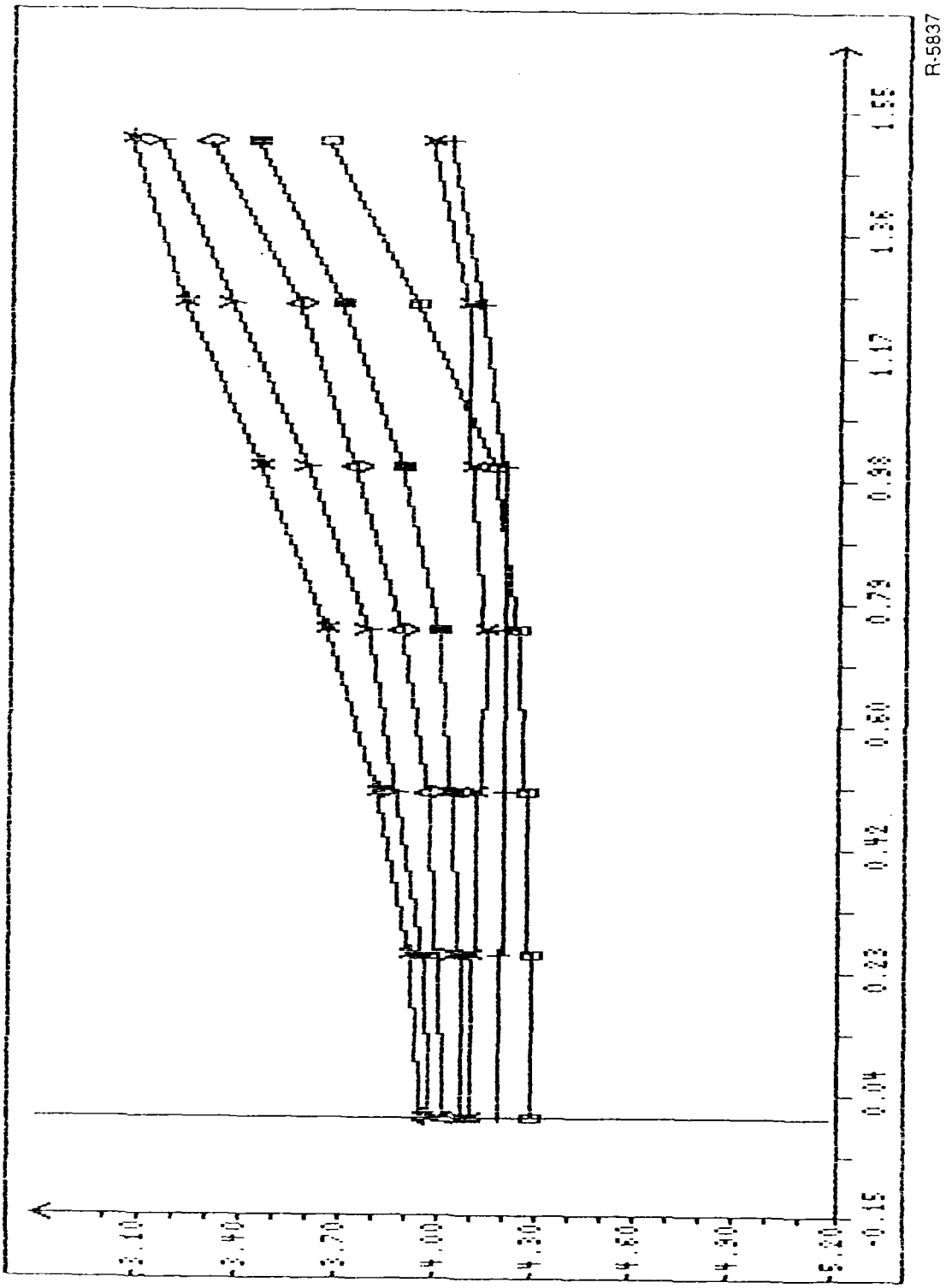


Figure 8-39. Plot of MSE vs. T(ε) for ε = .001 to .1 and g(ε) = .001

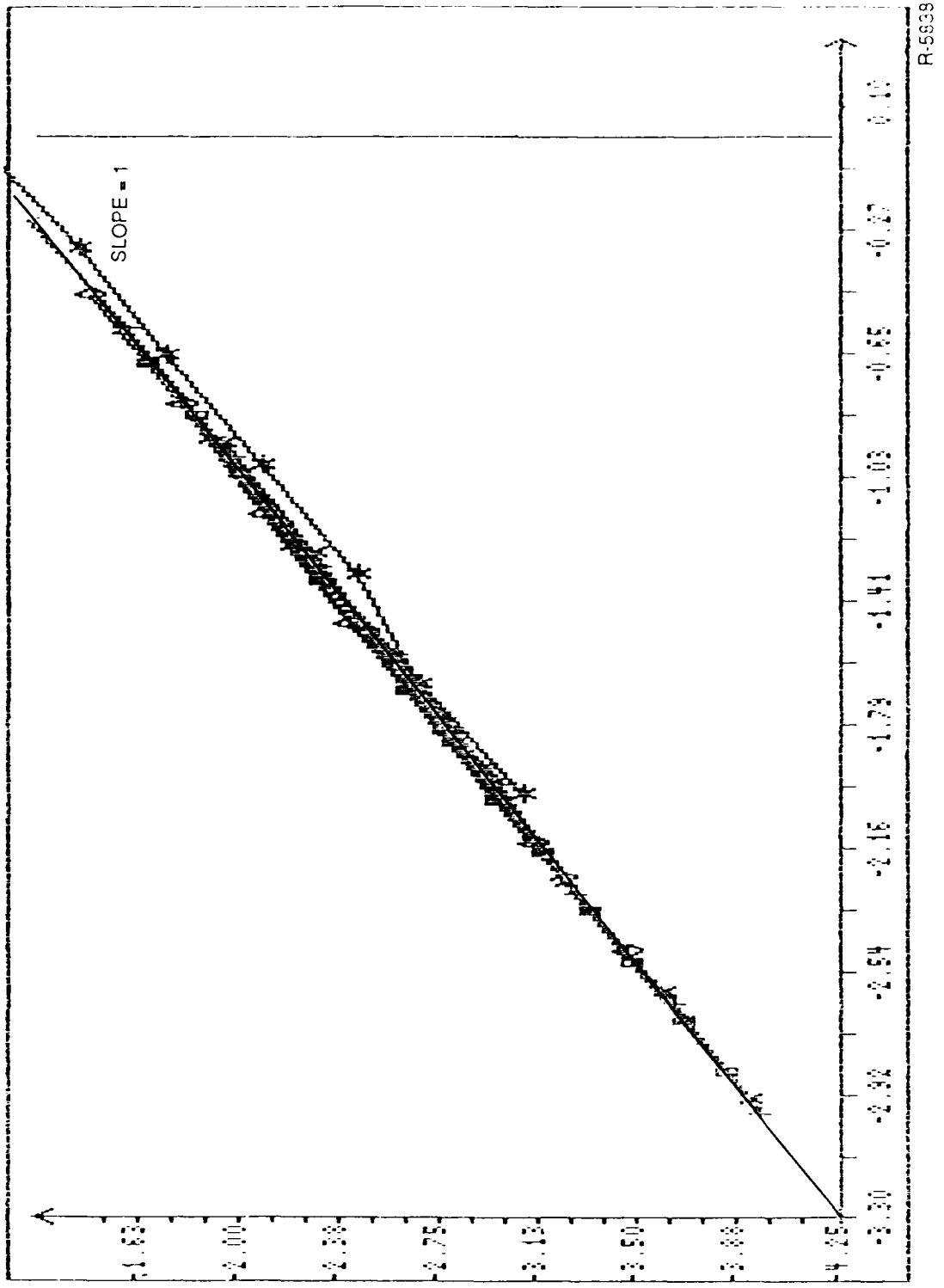


Figure 8-40. Plot of MSE vs. $\ell n(g(\epsilon)(1+\epsilon T^2(\epsilon)))^{1/2}$ for $\epsilon = .1$

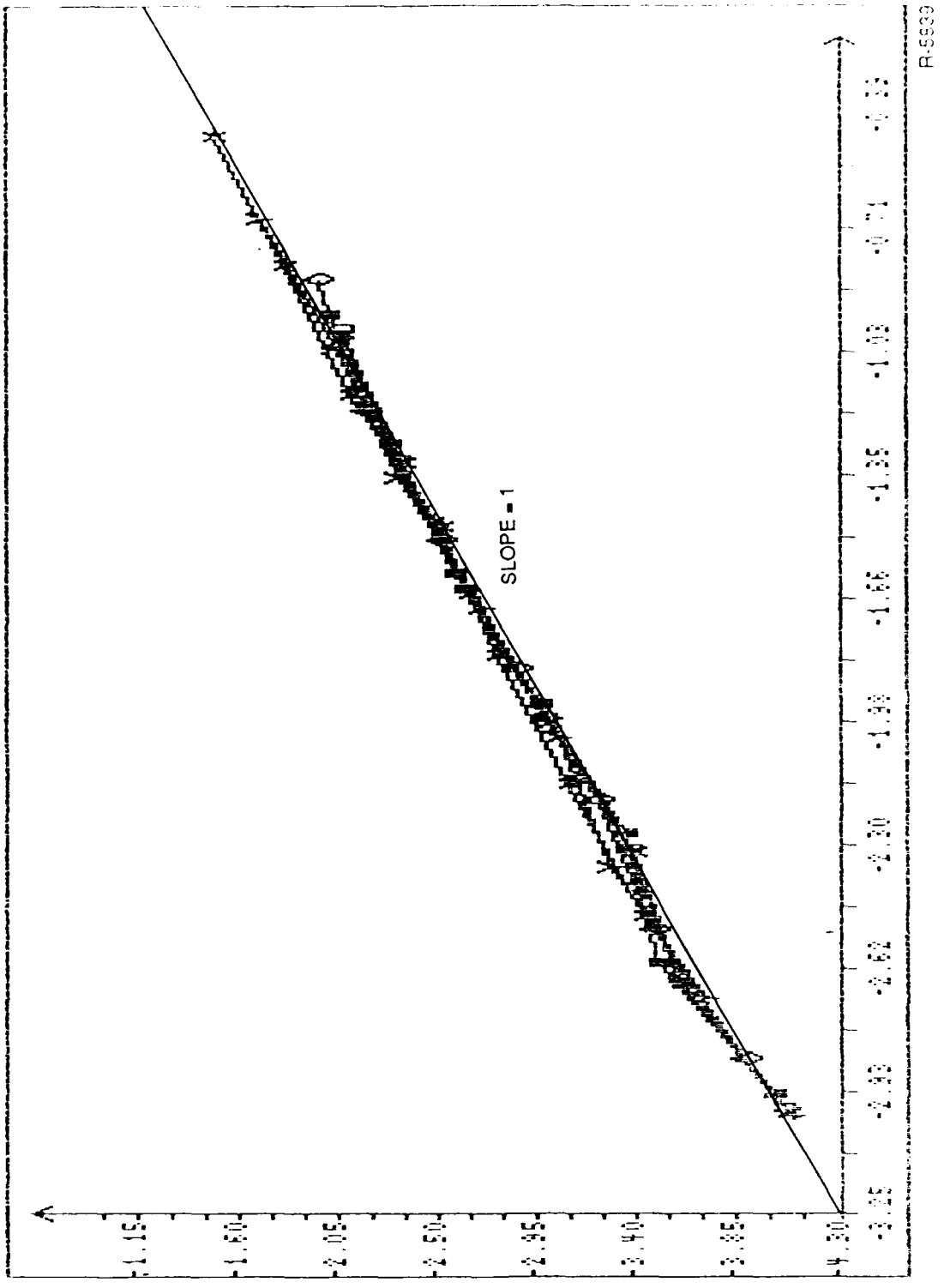


Figure 8-41. Plot of MSE vs. $\ln[g(\varepsilon)(1+\varepsilon T^2(\varepsilon))]^{1/2}$ for $\varepsilon = .01$

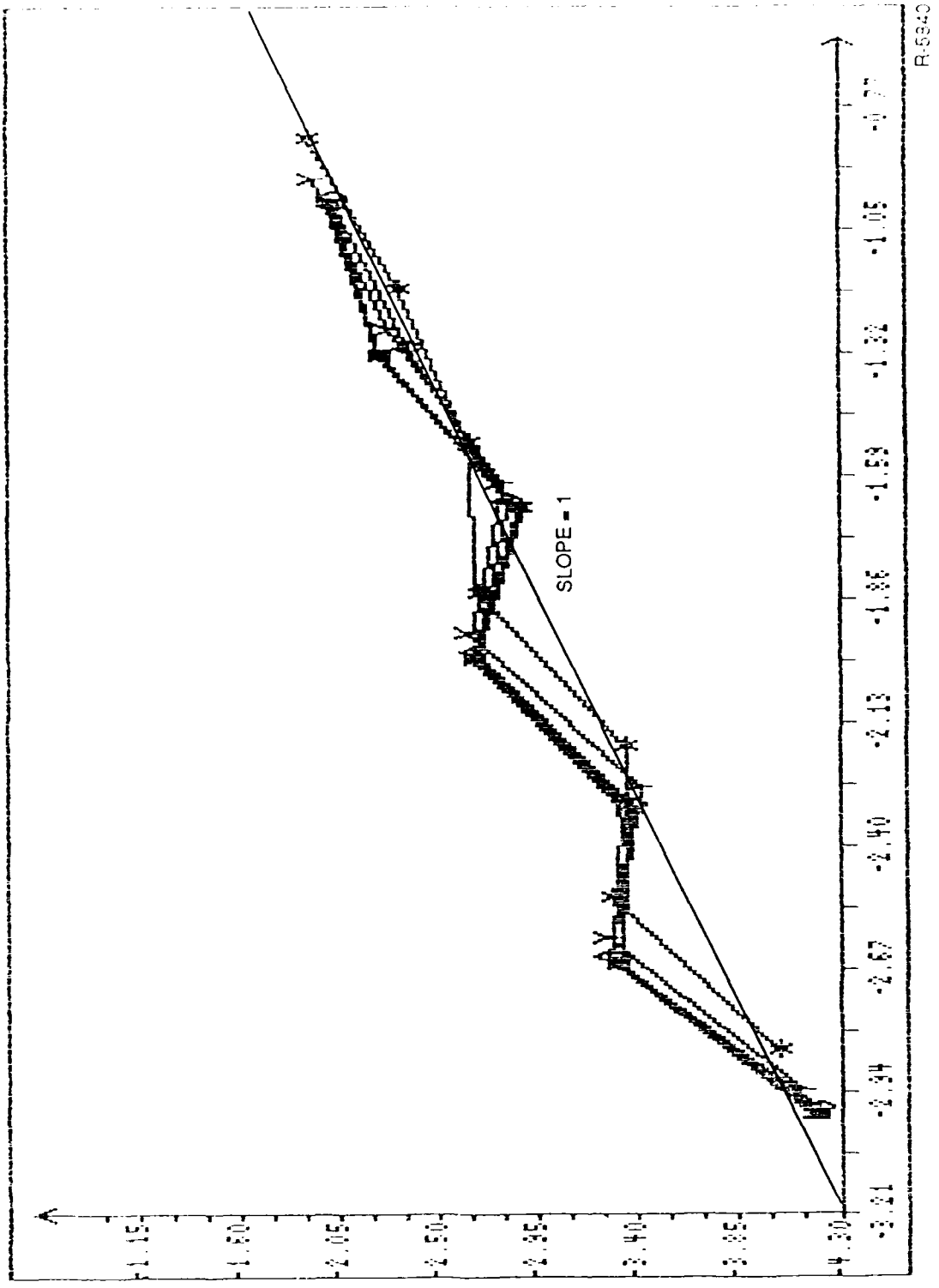


Figure 8-42. Plot of MSE vs. $\ln[g(\epsilon)(1+\epsilon T^2(\epsilon))]^{1/2}$ for $\epsilon = .001$

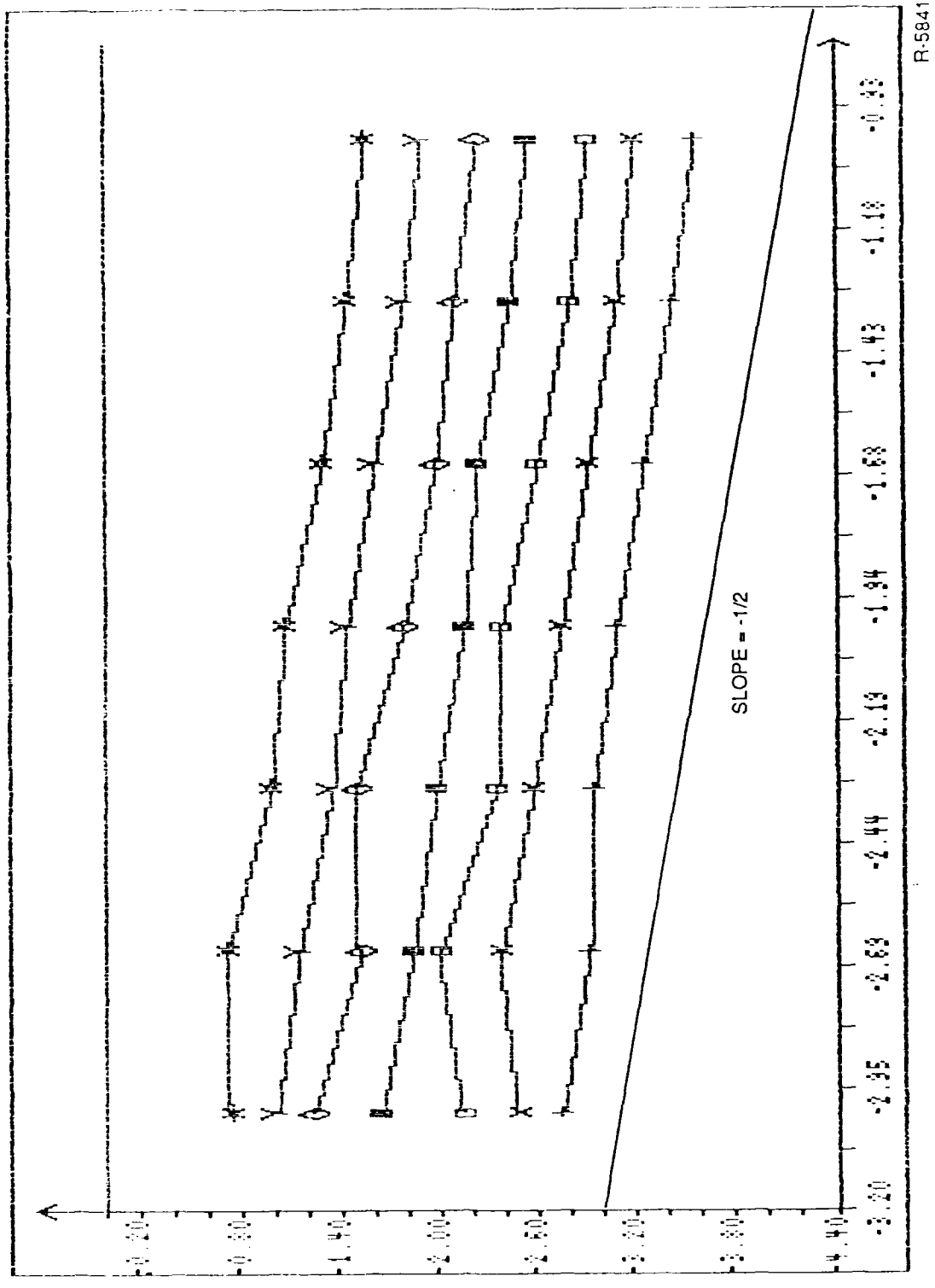


Figure 8-43. Plot of MSI vs. ε for g(ε) from .001 to .1 and T(ε) = 1

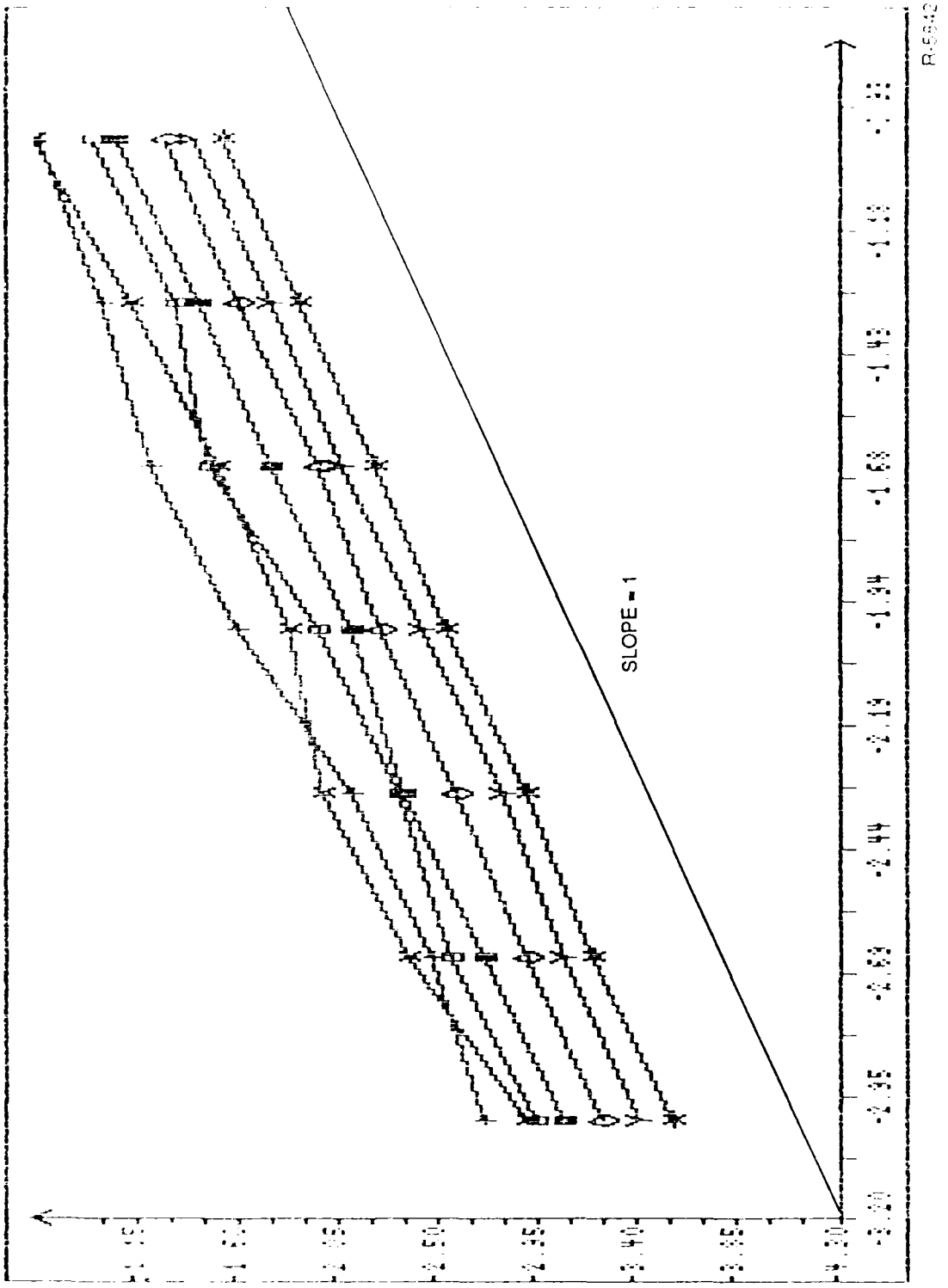


Figure 8-44. Plot of MSI vs. $g(\epsilon)$ for ϵ = .001 to .1 and $T(\epsilon)$ = 1

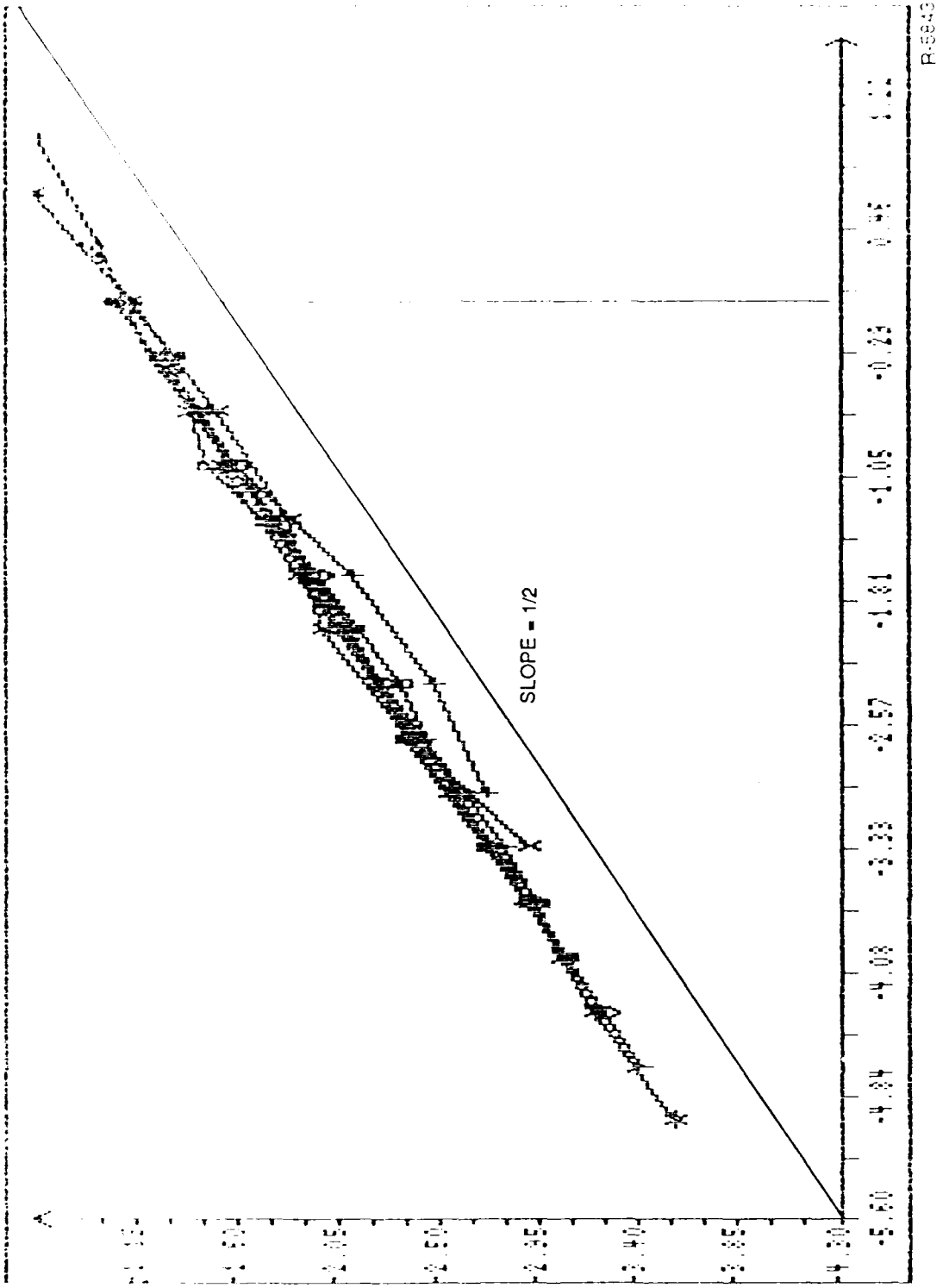


Figure 8-45. Plot of MSI vs. $\ln[g^2(\epsilon)/\epsilon]$ with $T(\epsilon) = 1$

nonlinear, which we believe is only the result of the time interval for simulation, 3000, being short relative to the average slow jump time (1000).

A conjecture regarding the mean squared value of "information" at the output of the filter as a function of $g(\varepsilon)$ and ε was formulated using Figs. 8-43 to 8-45. Figures 8-43 and 8-44 plot mean square information versus ε and $g(\varepsilon)$ with resulting slopes of -1/2 and 1. These two plots lead to the conjecture of Eq. 8-45.

$$\lim_{\substack{T\to\infty \\ \varepsilon\to 0}} \frac{1}{T} \int_0^T \left| P_L(t) - \frac{\gamma_2}{\gamma_1+\gamma_2} \right|^2 dt = K_3 \frac{g^2(\varepsilon)}{\varepsilon} \quad . \tag{8-45}$$

To provide support for this conjecture, mean square information was plotted against the expression in Eq. 8-45 with K_3 set to 1. The result was a linear plot (Fig. 8-45) with slop 1 providing strong support for the conjecture. If we combine Eqs. 8-44 and 8-45, we can obtain a condition for the relative error of the approximation to go to zero. First we divide the equations to obtain Eq. 8-46.

$$\lim_{\substack{T\to\infty \\ \varepsilon\to 0}} \frac{\int_0^T \left| P_L - p^L \right|^2 dt}{\int_0^T \left| P_L - \frac{\gamma_2}{\gamma_1+\gamma_2} \right|^2 dt} = K_4 \, \varepsilon\left(1+\varepsilon T^2(\varepsilon)\right) \quad . \tag{8-46}$$

The expression on the right side of Eq. 8-46 approaches 0 as ε approaches 0, if $T(\varepsilon)$ satisfies Eq. 8-47.

$$T(\varepsilon) = 0(\varepsilon^k) \quad , \quad k > -1 \quad . \tag{8-47}$$

In summary, this section provides conjectures for the conditions on the order of ε, $g(\varepsilon)$ and $T(\varepsilon)$ required to provide a nonzero quantity of information at the filter output and relatively unimportant error for the front-end/back-end approximation to the ideal filter.

8.4 MONTE CARLO CHARACTERIZATIONS OF BATCH STATISTICS L(t)

In Section 6, a quantity L(t) was introduced for the discrete measurement model as the likelihood ratio defined earlier by Eq. 6-64.

$$L(t) = \frac{\Pr\{x(\tau),\ \tau<t \mid \rho(t) \text{ on the right}\}}{\Pr\{x(\tau),\ \tau<t \mid \rho(t) \text{ on the left}\}} \quad . \tag{8-48}$$

This value L(t) is used to perform Bayesian updates in the back-end processor using Eq. 6-78. It would be useful to obtain an analytical approximation to the conditional distributions of L(t) in order to predict performance characteristics for the estimator structure. As a first step in obtaining such a formula, a series of simulations were performed to obtain histograms for the value of $\ell(T(\varepsilon))$, where $\ell(T(\varepsilon)) = \ell n\ (L(T(\varepsilon)))$ and $T(\varepsilon)$ is the time interval for each front-end batch. Figures 8-46 and 8-47 show normalized histograms for $\ell(T(\varepsilon))$ based on 200 sample paths and $T(\varepsilon) = 1$ and 3, while Figs. 8-48 through 8-50 show histograms for $T(\varepsilon) = 10$, 31, and 100 and using 2500 runs. Each figure actually plots two histograms, each normalized to unit area, one for $\rho(t)$ on the left and for for $\rho(t)$ on the right. The amount of separation between the curves provides a measure of the amount of information obtained regarding which pair of states the system is in (left or right).

The results for $T(\varepsilon) = 1$ and 3 are somewhat jagged, but are nearly on top of each other, and therefore a very small amount of information is obtained in

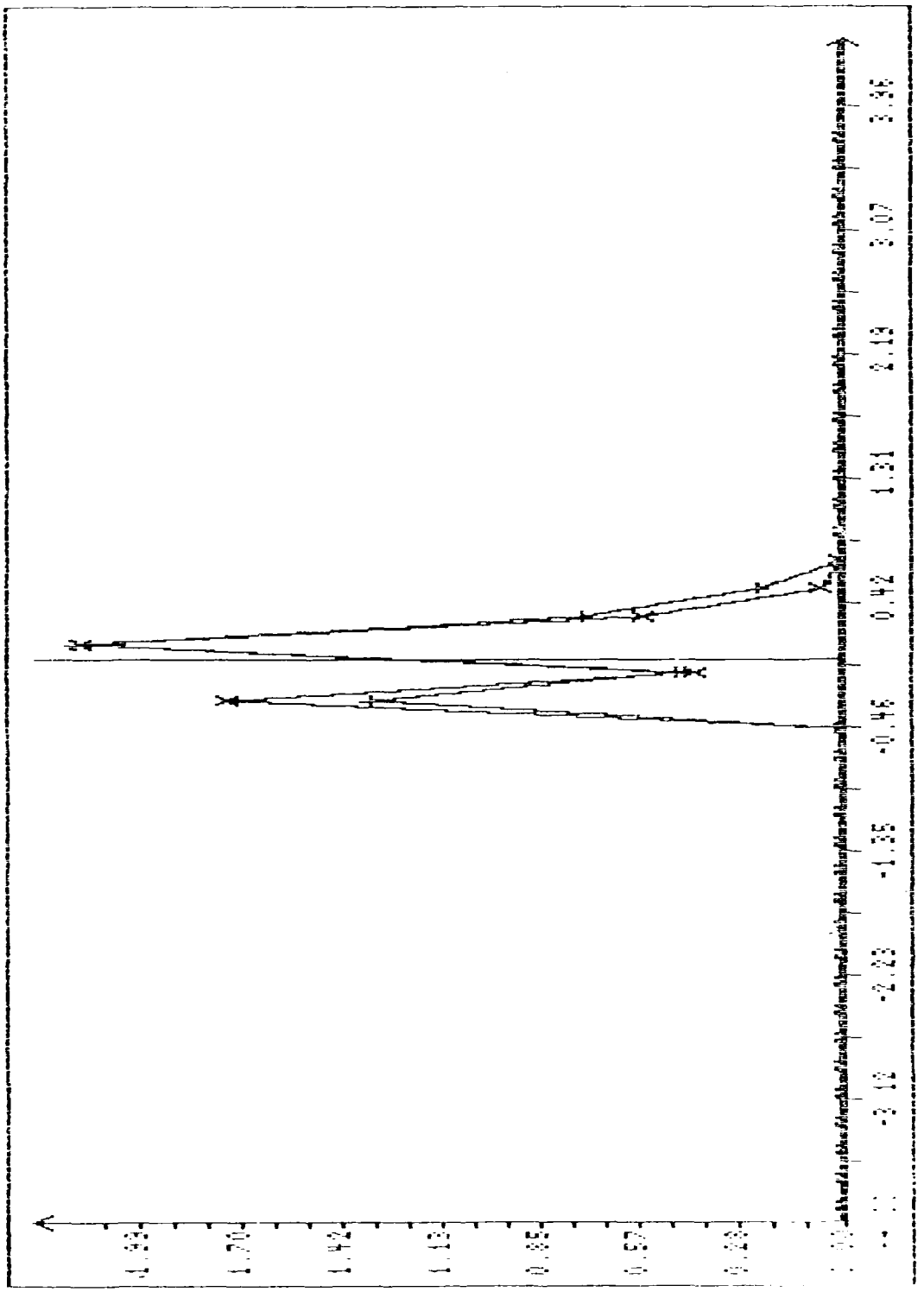
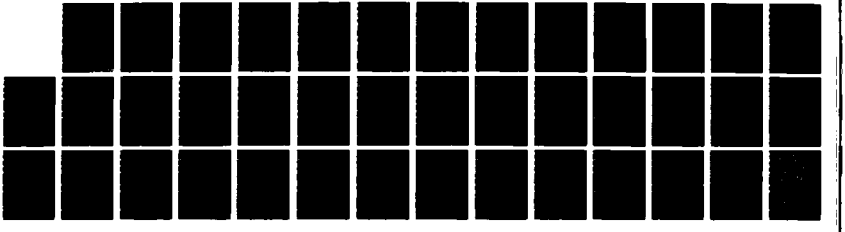
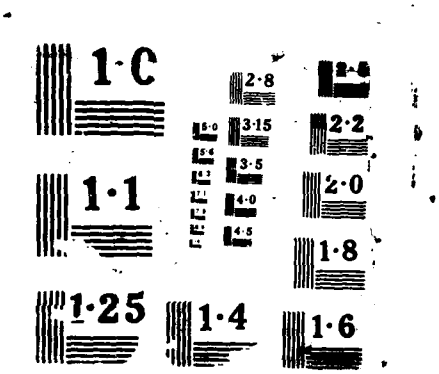


Figure 8-46. Plot of $p(\ell|L)$, $p(\ell|R)$ for $T = 1$, $g_1 = 0.3$ and 200 Runs

AD-A188 424 PASSIVE ACOUSTIC TRACKING AND FILTERING PROBLEMS WITH TIME SCALES(U) ALPHATECH INC BURLINGTON MA A CAROMICOLI ET AL NOV 87 TR-352 N00014-85-C-0349 3/3

UNCLASSIFIED F/G 17/1 NL




1·C
2·8
3·15
3·5
2·2
1·1
2·0
1·8
1·25
1·4
1·6

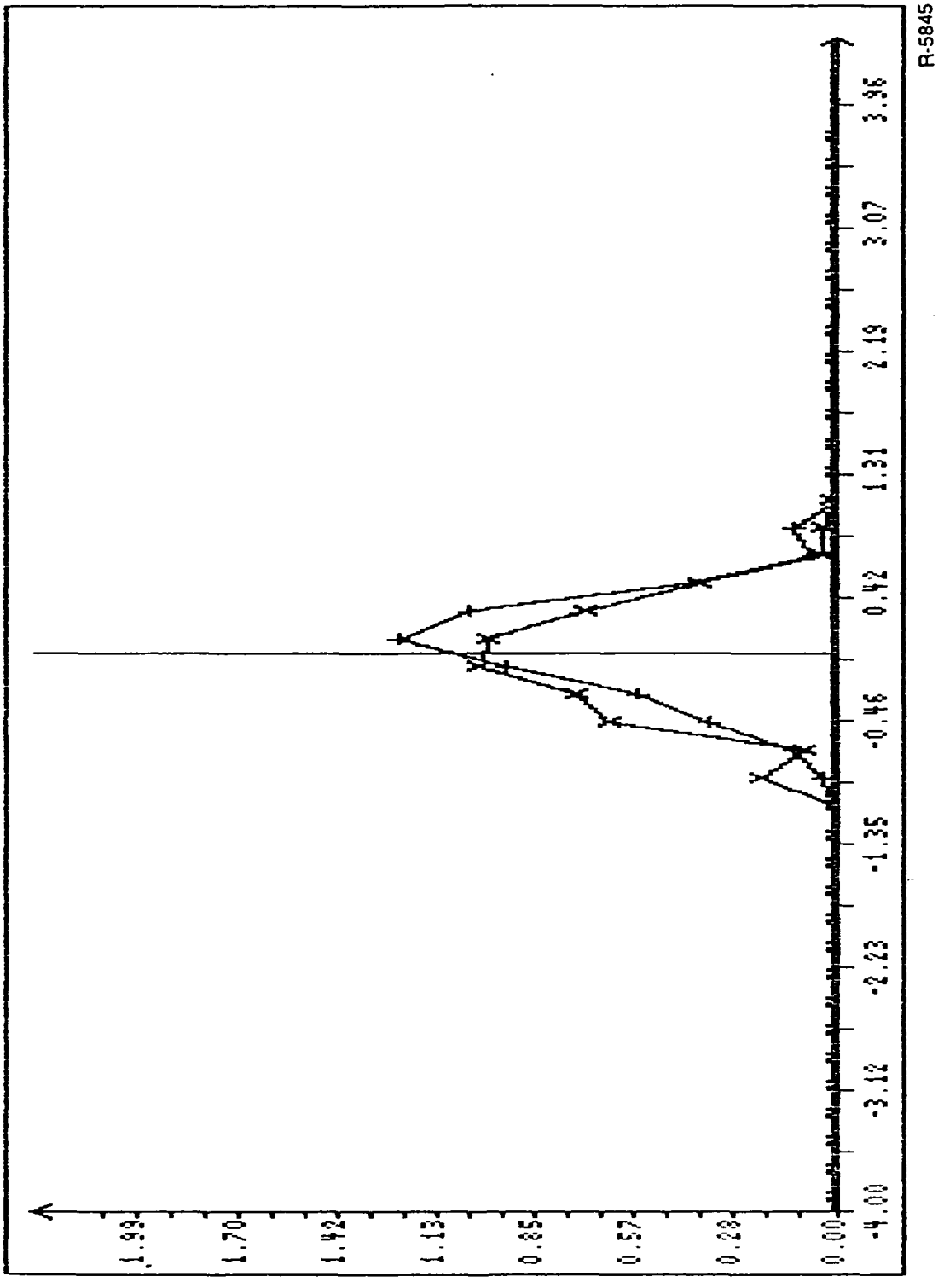


Figure 8-47. Plot of $p(\ell|L)$, $p(\ell|R)$ for T = 3, g_1 = 0.3 and 200 Runs

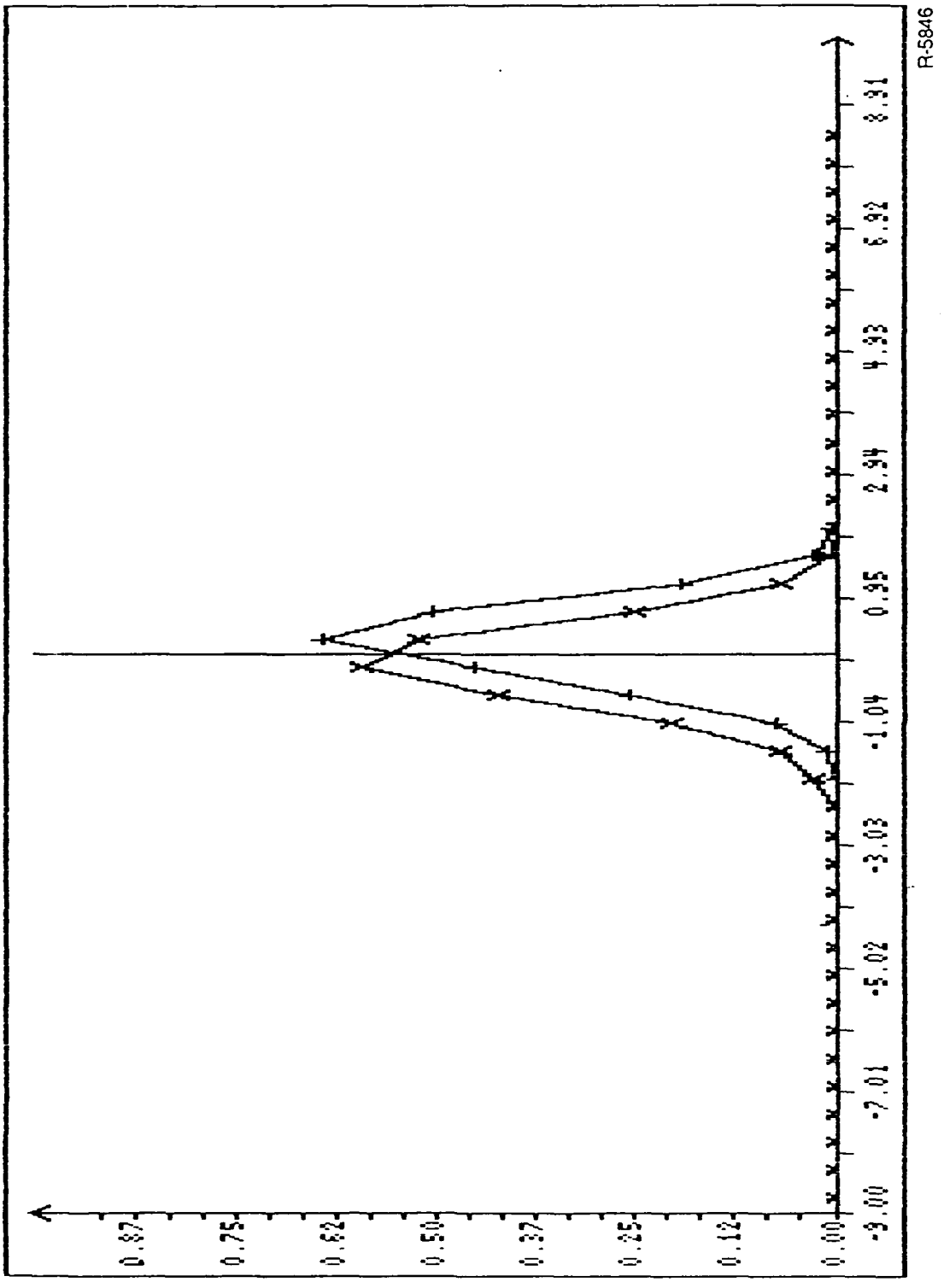


Figure 8-48. Plot of $p(\ell|L)$, $p(\ell|R)$ for T = 10, g_1 = 0.3 and 2500 Runs

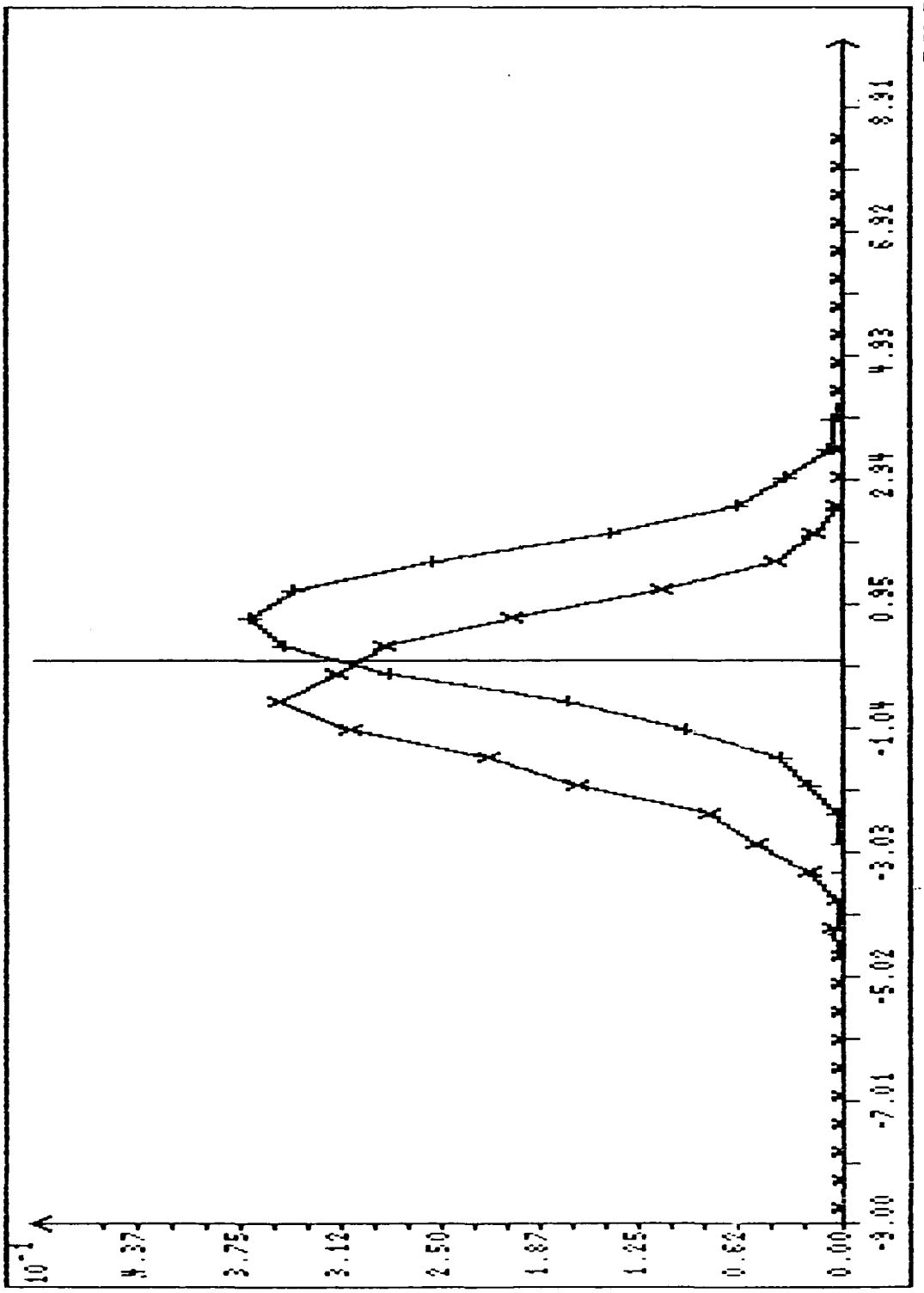


Figure 8-49. Plot of $p(\ell|L)$, $p(\ell|R)$ for T = 31, β_1 = 0.3 and 2500 Runs

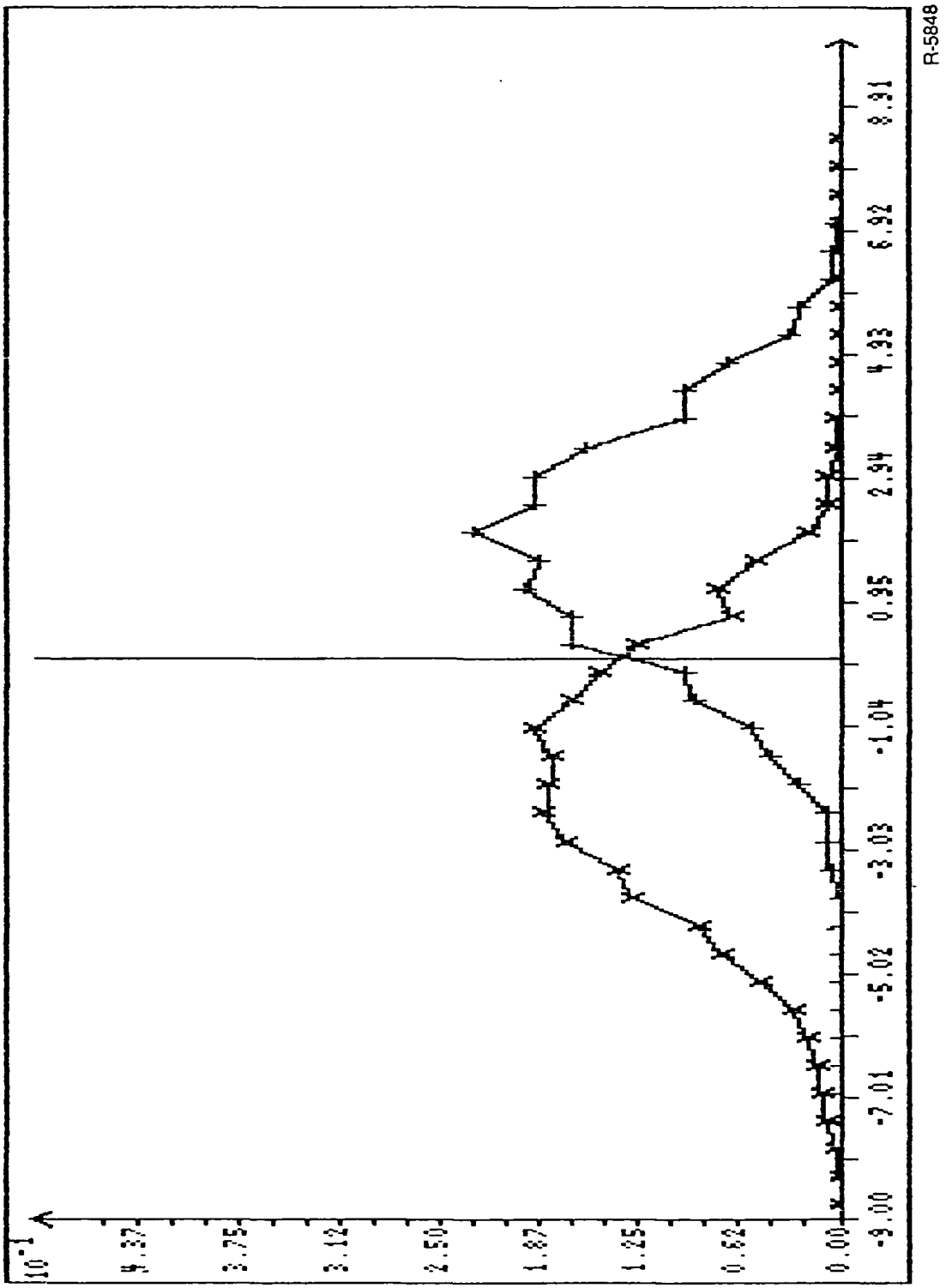


Figure 8-50. Plot of $p(\ell|L)$, $p(\ell|R)$ for T = 100, β_1 = 0.3 and 2500 Runs

1 or 3 time units when $g_1(\varepsilon) = 0.3$. The plots for $T(\varepsilon) = 10$, 31 and 100 show successively more separation between the peaks of the curves. With less overlap, there is much less chance of $\ell(t)$ or $L(t)$ providing "incorrect" information about the state to the back end.

The appearance of the histograms suggests that a Gaussian approximation to the conditional distributions might be effective. To obtain such an approximation, we need only calculate the mean and variance of $\ell(t)$. We start with the number of jumps K and the residence time in the the top state, assuming that the process is on the left. If we assume that $T(\varepsilon)$ is long enough for the fast dynamics to reach steady-state, we can calculate:

$$E[R_t(t)] = \frac{\lambda_2 t}{\lambda_1 + \lambda_2}$$

$$VAR[R_t(t)] = \frac{2\,\lambda_1\lambda_2 t}{(\lambda_1+\lambda_2)^3}$$

$$E[K(t)] = \frac{2\,\lambda_1\lambda_2 t}{(\lambda_1+\lambda_2)}$$

$$VAR[K(t)] = \frac{4\,\lambda_1\lambda_2(\lambda_1^2+\lambda_2^2)t}{(\lambda_1+\lambda_2)^3}$$

$$COV[R_t(t)\ K(t)] = \frac{2\,\lambda_1\lambda_2(\lambda_1-\lambda_2)t}{(\lambda_1+\lambda_2)^3}$$

Using these equations we can calculate the conditional mean and variance of ℓ given that we are on the left.

$$E[\ell|L] \cong C_1\, E[R_t|L] + \frac{1}{2}\, C_2\, E[K|L] - C_4 \qquad (8\text{-}49)$$

$$VAR[\ell|L] \cong C_1^2\, VAR[R_t|L] + C_2^2\, VAR[K|L] + 2\, C_1\, C_2\, COV[R_t K|L] \qquad (8\text{-}50)$$

where C_1, C_2, and C_4 are defined in the Appendix C (subsection C.2). Using Eqs. 8-49 and 8-50 and their counterparts if the process is on the right, the approximate conditional distributions for $\ell(t)$ were calculated and plotted for each value of $T(\varepsilon)$ that was used in the simulations. Comparing these approximations (Figs. 8-51 through 8-55) to the simulations we see there is close agreement.

8.5 DISCUSSION

The intention of the simulations was to provide evidence supporting both conjectures and theoretical results from Sections 6 and 7.

For the noise model we note some important features of the simulation results:

1. For large signal strengths, the estimator behaves in a switch-like fashion, detecting state transitions shortly after they occur.

2. For low signal strengths, insufficient information is obtained between transitions to determine the current state of the system.

3. All the approximate filters approach the performance of the optimal filter for appropriate parameter ranges.

4. The mean square approximation error of the aggregate model filter is a factor $O(\varepsilon)$ smaller than the mean square "information" in the output of the optimal filter for small ε and $g(\varepsilon) < \varepsilon^{1/2}$.

5. For $g(\varepsilon) > \varepsilon^{1/2}$, the mean square value of the information rate is relatively unaffected by decreases in the magnitude of ε.

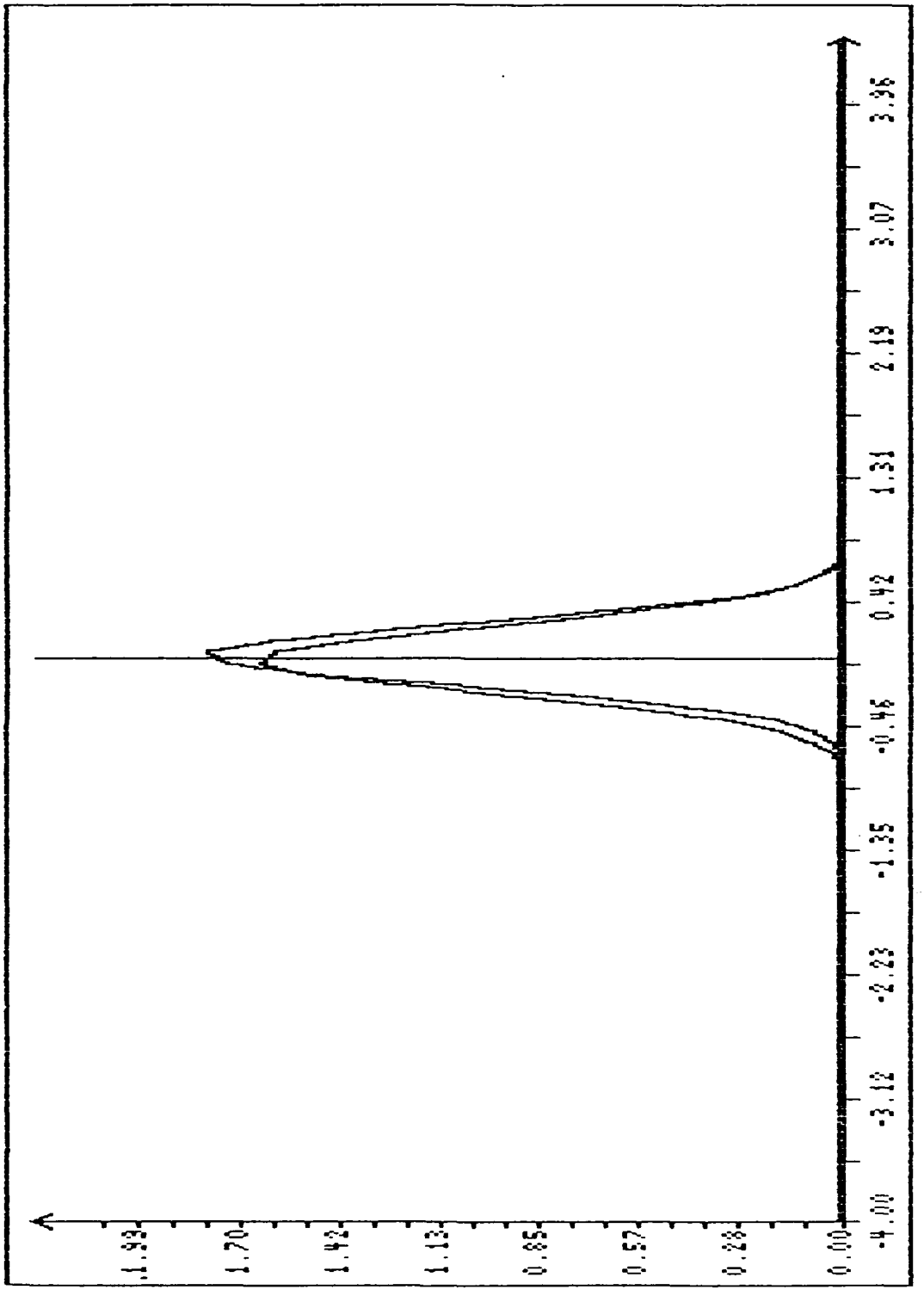


Figure 8-51. Gaussian Approximations for $g(\varepsilon) = 0.3$, $T(\varepsilon) = 1$

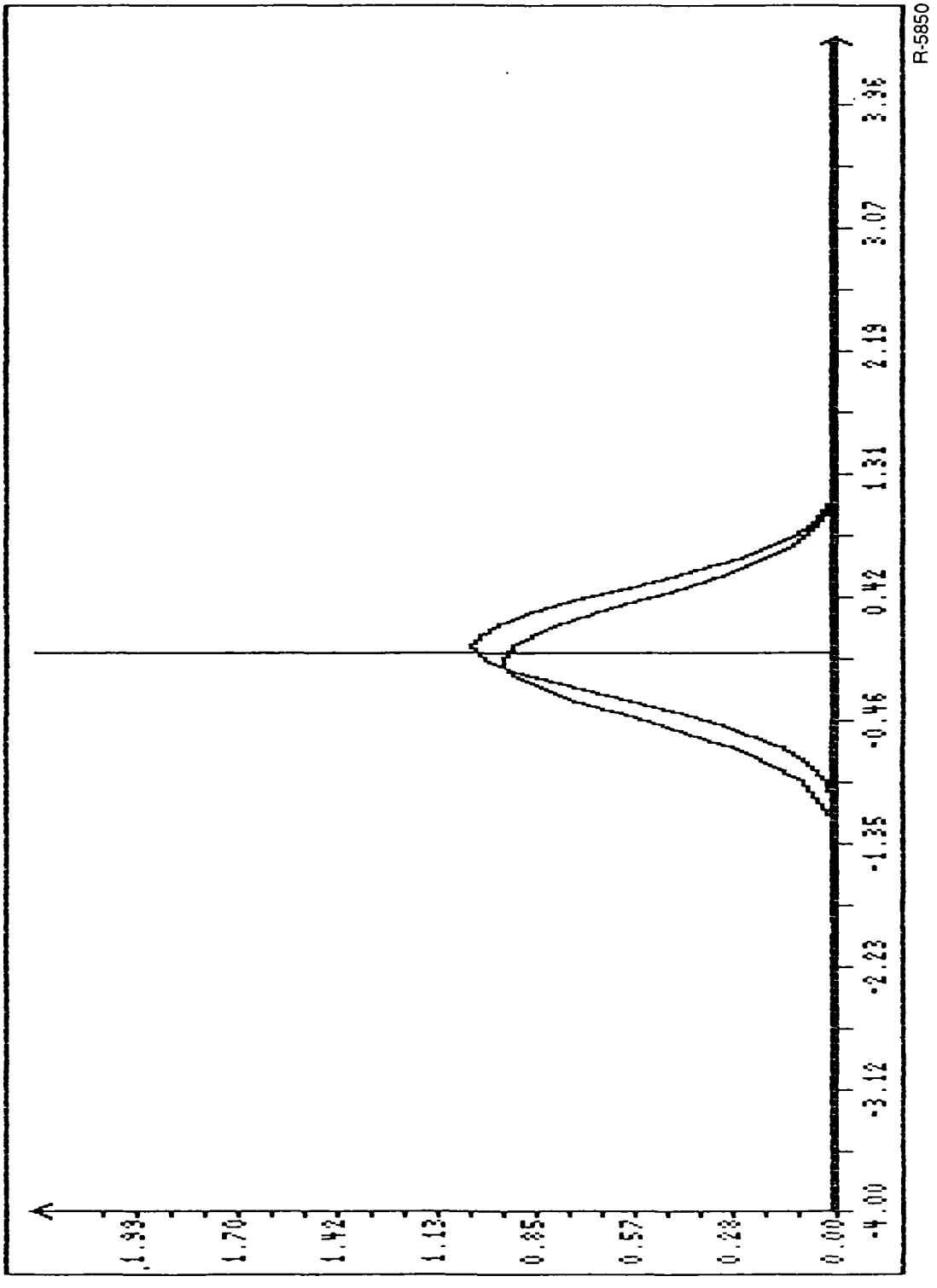


Figure 8-52. Gaussian Approximations for $g(\epsilon) = 0.3$, $T(\epsilon) = 3$

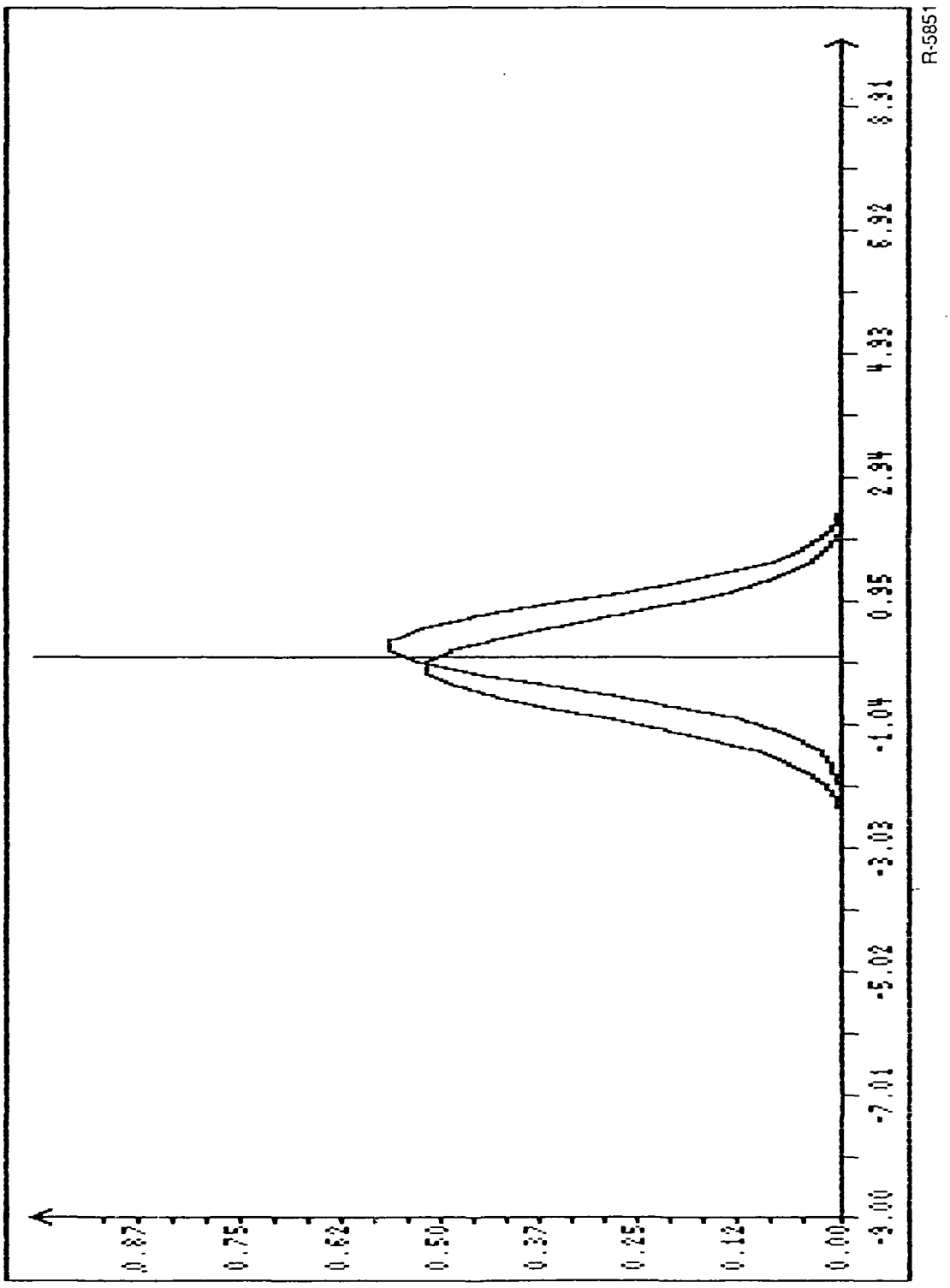


Figure 8-53. Gaussian Approximations for $g(\epsilon) = 0.3$, $T(\epsilon) = 10$

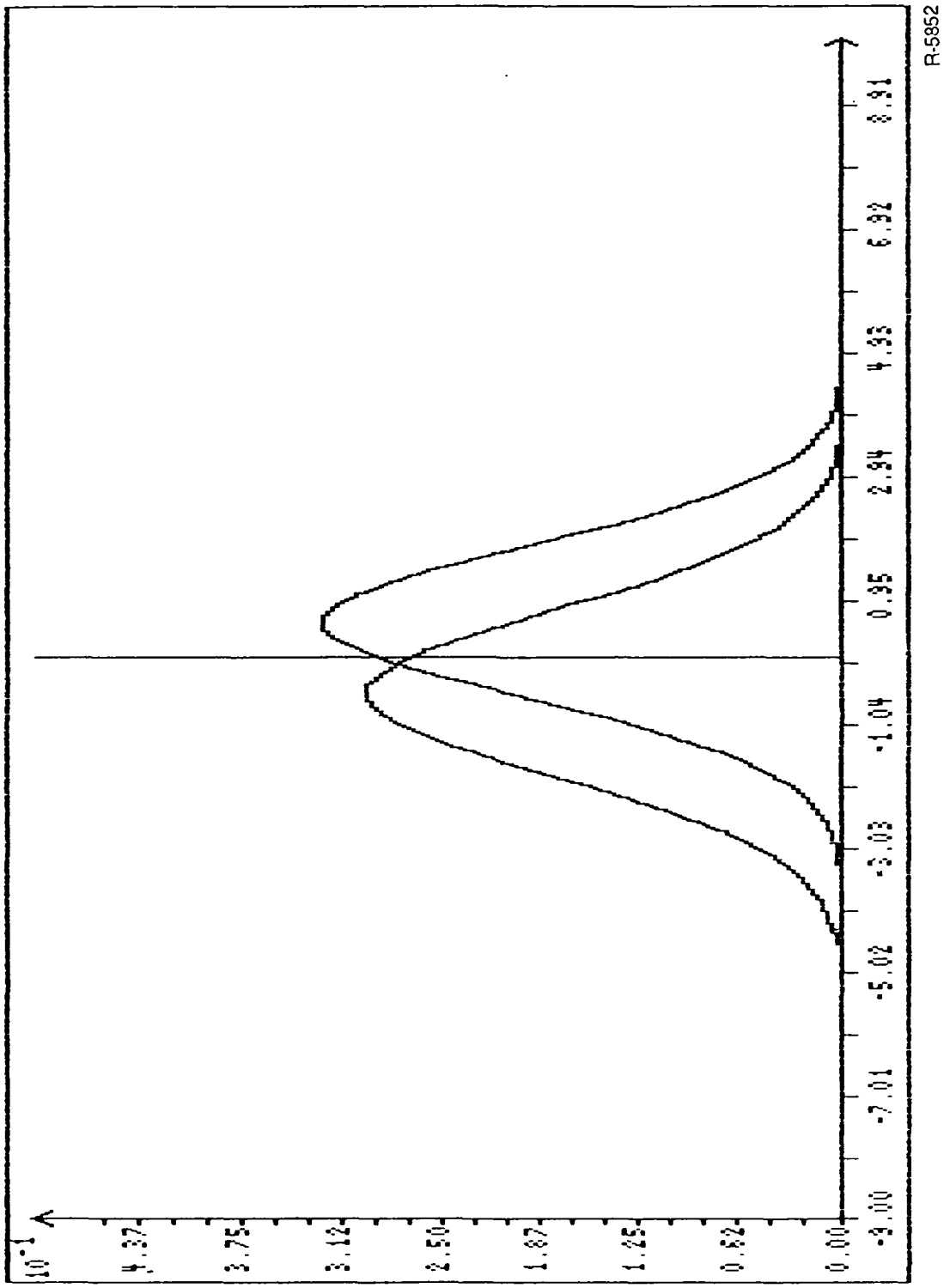


Figure 8-54. Gaussian Approximations for $g(\epsilon) = 0.3$, $T(\epsilon) = 31$

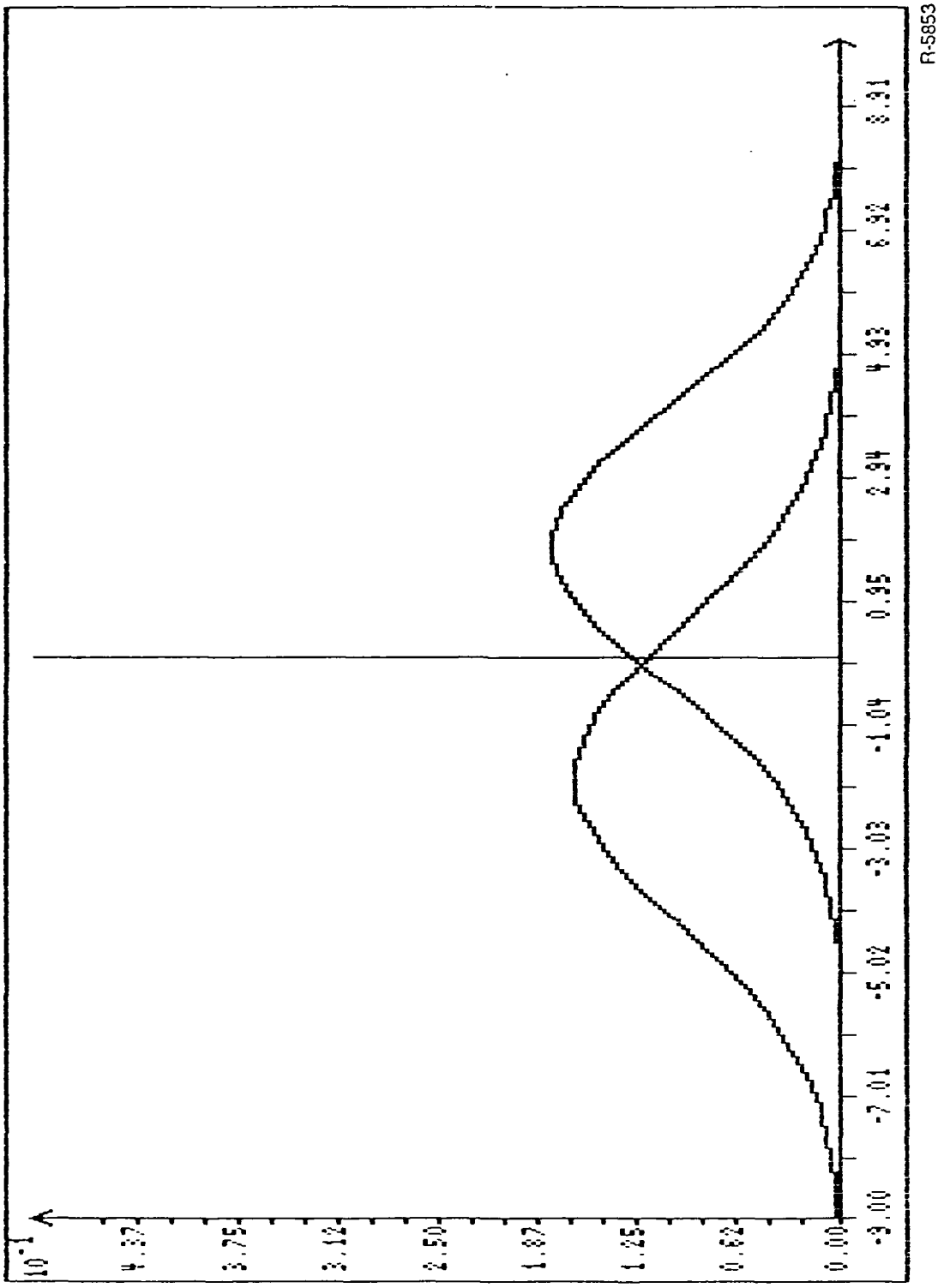


Figure 8-55. Gaussian Approximations for $g(\epsilon) = 0.3$, $T(\epsilon) = 100$

For the discrete measurement model we can make the following general comments:

1. The choice of the magnitude of $g_i(\varepsilon)$, the differences in the fast transition rates, has a critical effort on filter performance which can vary from providing virtually no information ($g_i(\varepsilon)$ very small) to switch-like performance ($g_i(\varepsilon)$ large).

2. For small values of $g(\varepsilon)T(\varepsilon)$ a differential equation on the slow time scale can be used to approximate the optimal filter with good accuracy.

3. The optimal filter can be approximated by a two-stage structure with front end calculating R_t and K and back end carrying out prediction and update steps given appropriate conditions on the system and filter parameters. This condition is conjected to be that the expression in Eq. 8-46 be small.

4. The conditional distribution of the statistic $\ell(t)$, the log of the likelihood ratio defined by Eq. 6-64 can be well-approximated by a Gaussian distribution. Furthermore, the mean and variance of the distribution can be approximated by simple functions of the fast transition rates.

SECTION 9

CONCLUSIONS

In this report we have described the results of our research on alternate architectures for nonlinear filtering problems possessing the key features found in ASW problems. Specifically, we have focused on systems in which a slowly-varying part of the state influences or modulates the behavior of a much more quickly varying portion of the state. This fast variable is then observed, and the ultimate objective is to estimate the slow variables.

As we have discussed, this problem can be cast in the framework of perturbation methods. In Section 4 we provide a survey of research on perturbed estimation problems and discuss the architectural implications of these various results. A conclusion of this survey is that there are no previously-developed results that are applicable to problems of the type of interest in this study.

There are several key ideas that we wished to address in this study. One was to develop an understanding of the relationship between the time scale of the slow process and the order of the magnitude of the measurement signal-to noise ratio. The second was to investigate alternate suboptimal architectures and in particular the so-called front-end/back-end architecture commonly used in ASW systems. In such a system, the front end processes a batch of data to produce an estimate of the slow variables based on the assumption that the slow variables are constant over the batch. This sequence of batch-produced

estimates is then fed into a back-end processor that tracks the slow variations in the slow process. Questions of interest are: when is this architecture a good one and how long should the batches be when compared to the natural scales determined by the slow process and by the measurement information rate.

Our investigation of these and several related questions has been carried out using a relatively simple, four-state Markov process possessing two time scales. We have used two different measurement models in our study: one in which poor quality, noise-corrupted measurements of the fast process are available and one in which we have perfect measurements of the fast process but the influence of the slow process on the fast one is quite weak. In this context we have been able to explore at great depth a variety of suboptimal filter structures and have obtained both theoretical results and convincing evidence from simulations concerning the asymptotic optimality of several suboptimal estimators.

In particular, for the problem with poor noise-corrupted estimates:

1. We have proven an important result showing the asymptotic optimality of a suboptimal slow estimator that essentially averages out the fluctuations in the measurements due to the fast variables. This result provides a clear picture of the relationship between time scales and the asymptotic quality of the measurements. These results, and several stronger sample path properties, have been demonstrated in our simulations.

2. We have implemented front-end/back-end structures both using the full, four-state model and the same averaging strategy described in 1. Simulations have shown that both of these estimators achieve asymptotically the same performance as the optimal filter. In addition, these results suggest particular asymptotic formulas for the length of the time interval used in the front-end batch processor.

For the case perfect fast measurements but weak slow-to-fast coupling:

1. We have demonstrated via simulations that an extremely slow integration of the optimal estimation equations -- which is equivalent to a simple front-end counter and interruptable clock followed by a slow back-end processor -- performs asymptotically as well as the optimal.

2. We have also studied the standard front-end/back-end algorithms for this problem. Our extensive simulations provide extremely convincing evidence supporting our conjectures concerning asymptotic optimality and the relationships among the orders of magnitude of time scales, slow-to-fast couplings, and front-end batch lengths.

The results presented in this report not only are of significance in their own right, but they also provide clear direction for further work: in the theoretical verification of the numerous conjectures developed through our simulations; in developing higher-order approximations that indicate the value of slightly more complex estimators; and in extending these ideas to continuous-state nonlinear systems.

REFERENCES

1. Knight, W.C., R.G. Pridham, and S.M. Kay, "Digital Signal Processing for Sonar," Proceedings of the IEEE, Vol. 69, November 1981, pp. 1451-1507.

2. Bucy, R.S., J.M.F. Moura, and A.J. Mallinckrodt, "A Monte Carlo Study of Absolute Phase Determination," IEEE Transactions on Information Theory, Vol. IT-29, July 1983, pp. 509-520.

3. Kushner, H.J., Approximation and Weak Convergence Methods for Random Processes, MIT Press, Cambridge, MA, 1984.

4. Urick, R.J., Sound Propagation in the Sea, Peninsula Publishing, Los Altos, CA, 1982.

5. McKean, H.P., Stochastic Integration, Academic Press, New York, 1969.

6. Washburn, R.B., "Unified Theory for Airborne Acoustic ASW Processing and Tracking," ALPHATECH Technical Report TR-161, 1983.

7. Washburn, R.B., D. Teneketzis, and A.S. Willsky, "Performance Analysis of Signal Processing and Tracking Systems for Airborne Acoustic ASW," ALPHATECH Technical Report TR-206, July 1984.

8. Jazwinski, A.H., Stochastic Processes and Filtering Theory, Academic Press, NY, 1970.

9. Kokotovic, P.V., R.E. O'Malley, Jr., and P. Sannuti, "Singular Perturbations and Order Reduction in Control Theory - An Overview," Automatica, Vol. 12, March 1976, pp. 123-132.

10. Saksena, V.R., J. O'Reilly, and P.V. Kokotovic, "Singular Perturbations and Time-Scale Methods in Control Theory: Survey 1976-1983," Automatica, Vol. 20, May 1984, pp. 273-293.

11. Davis, M.H.A. and S.I. Marcus, "An Introduction to Nonlinear Filtering," in Stochastic Systems: The Mathematics of Filtering and Identification and Applications, pp. 53-75, edited by M. Hazewinkel and S. Mitter, Reidel Publishing Company, 1981.

12. Lipster and A.N. Shiryayev, Statistics of Random Processes, Springer-Verlag, NY, 1977.

13. Wonham, W.M., "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering," SIAM Journal of Control, Vol. 2, 1965, pp. 347-369.

14. Willems, J.C., A. Kitapci, and L.M. Silverman, "Singular Optimal Control: A Geometric Approach," SIAM Journal of Control and Optimization, Vol. 24, March 1986, pp. 323-337.

15. Schumacher, J.M., "A Geometric Approach to the Singular Filtering Problem," IEEE Transactions on Automatic Control, Vol. AC-30, November 1985, pp. 1075-1083.

16. Shaked, U., "Explicit Solution to the Singular Discrete Time Stationary Linear Filtering Problem," IEEE Transactions on Automatic Control.

17. Krener, A.J., "The Asymptotic Approximation of Nonlinear Filters by Linear Filters," preprint.

18. Sannuti, P. and H.S. Wason, "Multiple Time-Scale Decomposition in Cheap Control Problems -- Singular Control," IEEE Transactions on Automatic Control, Vol. AC-30, July 1985, pp. 633-644.

19. Sannuti, P., "Direct Singular Perturbations of High-Gain and Cheap Control Problems," Automatica, Vol. 19, January 1983, pp. 41-51.

20. Saberi, A. and P. Sannuti, "Cheap and Singular Controls for Linear-Quadratic Regulators," IEEE Transactions on Automatic Control, Vol. AC-32, March 1987, pp. 208-219.

21. Saberi, A. and P. Sannuti, "Cheap control of a Linear Uniform Rank System Design by Composite Control," Automatica, Vol. 22, June 1986, pp. 757-759.

22. Sannuti, P. and H. Wason, "A Singular Perturbations Canonical System of Invertible Systems: Determination of Multivariable Root-Loci," International Journal of Control, Vol. 37, June 1983, pp. 1259-1286.

23. Bryson, A. and Y.-T. Ho, Applied Optimal Control, Hemisphere Publishing, Toronto, 1975.

24. Katzur, R., B.Z. Bobrovsky, Z. Schuss, "Asymptotic Analysis of the Optimal Filtering Problem for One-Dimensional Diffusions Measured in a Low-Noise Channel, Part I," SIAM Journal of Applied Mathematics, Vol. 44, June 1984, pp. 591-604.

25. Chow, J.H. and P. Kokotovic, "A Two-Stage Lyapunov-Bellman Feedback Design of a Class of Nonlinear Systems," IEEE Transactions on Automatic Control, Vol. AC-26, June 1981, pp. 656-663.

26. Haddad, A.H., "Linear Filtering of Singularly Perturbed Systems," IEEE Transactions on Automatic Control, Vol. AC-21, August 1976, pp. 515-519.

27. Marchetti, C., "Regular Perturbations in the Filtering of Markov Chains," preprint.

28. Marchetti, C., "Singular Perturbations in Filtering Problem of the Markov Chains," preprint.

29. Coderch, M., A.S. Willsky, S.S. Sastry, and D.A. Castanon, "Hierarchical Aggregation of Singuilarly Perturbed Finite State Markov Processes," Stochastics, Vol. 8, 1983, pp. 259-289.

APPENDIX A

SOLUTION OF THE WAVE EQUATION

Pressure in a three-dimensional homogeneous fluid with sound speed c satisfies the wave equation

$$\frac{1}{c^2} \cdot \frac{\partial^2 u}{\partial t^2} - \nabla^2 u = f(x,t) \qquad \text{(A-1)}$$

where $f(x,t)$ is the source field. We are interested in the case of a moving point source at $x_T(t)$ for which $f(x,t)$ is given by

$$f(x,t) = 4\pi\delta\left(x - x_T(t)\right) \cdot y_T(t) \qquad \text{(A-2)}$$

where δ is the three-dimensional Dirac delta function and $y_T(t)$ is the source signal. The solution of Eq. A-1 can be expressed in terms of Greens function as an integral

$$u(x,t) = \int \frac{f\left(\xi, t - \frac{|x-\xi|}{c}\right)}{4\pi \cdot |x-\xi|} d\xi \quad . \qquad \text{(A-3)}$$

Substituting Eq. A-2 in Eq. A-3 gives

$$u(x,t) = \int \frac{\delta\left(\xi - x_T\left(t - \frac{|x-\xi|}{c}\right)\right) \cdot y_T\left(t - \frac{|x-\xi|}{c}\right)}{|x-\xi|} d\xi \quad . \qquad \text{(A-4)}$$

To integrate Eq. A-4 we need to change variables from ξ to ζ defined by

$$\zeta = \xi - x_T\left(t - \frac{|x-\xi|}{c}\right) . \tag{A-5}$$

Note that the Jacobian of the transformation is

$$\left(\frac{\partial \zeta}{\partial \xi}\right) = I_3 - \frac{\dot{x}_T}{c}\frac{(x-\xi)^T}{|x-\xi|} \tag{A-6}$$

where the superscript T in Eq. A-6 denotes vector transposition and I_3 denotes the 3×3 identity matrix. The determinant of the Jacobian is needed to transform variables in the integral and is given by

$$\det\left(\frac{\partial \zeta}{\partial \xi}\right) = 1 - \frac{\langle \dot{x}_T, x-\xi\rangle}{c\cdot|x-\xi|} . \tag{A-7}$$

The integral in Eq. A-4 becomes

$$\int \left[1 - \frac{\langle \dot{x}_T, x-\xi\rangle}{c\cdot|x-\xi|}\right]^{-1} \frac{y\left(t - \frac{|x-\xi|}{c}\right)}{|x-\xi|}\,\delta(\zeta)\,d\zeta \tag{A-8}$$

where ξ in Eq. A-8 satisfies Eq. A-5. Thus, we get

$$u(x,t) = \left[1 - \frac{\langle \dot{x}_T, x-\xi\rangle}{c\cdot|x-\xi|}\right]^{-1} \frac{y\left(t - \frac{|x-\xi|}{c}\right)}{|x-\xi|} \tag{A-9}$$

where ξ is chosen to satisfy

$$0 = \xi - x_T\left(t - \frac{|x-\xi|}{c}\right) . \tag{A-10}$$

Letting

$$\tau = \frac{|x-\xi|}{c} \quad , \tag{A-11}$$

we see that

$$c\tau = |x-x_T(t-\tau)| \quad , \tag{A-12}$$

$$1 - \dot{\tau} = \left[1 - \frac{\langle \dot{x}_T, x-\xi \rangle}{c \cdot |x-\xi|} \right]^{-1} \quad , \tag{A-13}$$

and finally

$$u(x,t) = \frac{y(t-\tau)}{c\tau} (1-\dot{\tau}) \quad . \tag{A-14}$$

APPENDIX B

FILTERING EQUATIONS FOR PARTIALLY-OBSERVED FINITE-STATE CONTINUOUS-TIME MARKOV PROCESS

B.1 PROBLEM FORMULATION AND FILTER EQUATIONS

Suppose that $x(t)$ is a finite-state continuous-time Markov process and define the observation process $y(t)$ by

$$y(t) = h\big(x(t)\big) \quad . \tag{B-1}$$

We will derive a stochastic differential equation for the conditional probability $\pi_t(\xi)$ defined by

$$\pi_t(\xi) = \Pr\{x(t)=\xi \,|\, y(\tau), 0 \geq \tau \geq t\} \quad . \tag{B-2}$$

Let $\lambda(\xi|\xi')$ denote the transition rate from ξ' to ξ. That is,

$$\Pr\{x(t+\Delta)=\xi \,|\, x(t)=\xi'\} = \delta_{\xi\xi'} + \Delta\cdot\lambda(\xi|\xi') + o(\Delta) \quad , \tag{B-3}$$

where $\Delta>0$, $\delta_{\xi\xi'}$ is the Kronecker delta function, and $o(\Delta)$ is an error term that tends to 0 faster than Δ as Δ tends to 0.

Assume that $x(t)$ is a version of the Markov process which has right continuous sample paths and which has, at most, a finite number of discontinuities in any finite time interval. Let J_t denote the counting process associated with y. That is, J_t is the number of discontinuities of y in the interval $[0,t]$.

We will show that $\pi_t(\xi)$ satisfies the stochastic differential equation

$$\pi_t(\xi) = \pi_0(\xi) + \int_0^t F_s \, ds + \int_0^t G_s \, dJ_s \tag{B-4}$$

where the first integral is Lebesgue and the second is Stieltjes. If τ_k, $1<k$, are the transition times of the y process (and therefore also of the counting process J), then the Stieljes integral in this case is simply

$$\int_0^t G_s \, dJ_s = \sum_{\tau_k \leq t} G_{\tau_k} \quad . \tag{B-5}$$

The integrands F_t and G_t are defined as follows.

$$F_t = \delta_{h(\xi)y(t)} \cdot \left[\sum_{\xi'} \lambda(\xi|\xi')\pi_t(\xi') - \pi_t(\xi) \cdot \sum_{\xi'} \lambda(y(t)|\xi')\pi_t(\xi') \right] \tag{B-6}$$

where

$$\lambda(y(t)|\xi') = \sum_{h(\xi'')=y(t)} \lambda(\xi''|\xi') \quad , \tag{B-7}$$

$$G_t = \delta_{h(\xi)y(t)} \cdot \frac{\sum_{\xi'} \lambda(\xi|\xi')\pi_{t-}(\xi')}{\sum_{\xi'} \lambda(y(t)|\xi')\pi_{t-}(\xi')} - \pi_{t-}(\xi) \tag{B-8}$$

where

$$\pi_{t-}(\xi) = \lim_{s \uparrow t} \pi_s(\xi) \quad . \tag{B-9}$$

A special case of this general result is when x(t) has two components

$$x(t) = \left(x_1(t), x_2(t)\right) \tag{B-10}$$

and

$$y(t) = h\big(x(t)\big) = x_2(t) \quad . \tag{B-11}$$

In this case the stochastic differential equation for

$$\pi_t(\xi_1) = \Pr\{x_1(t)=\xi \,|\, y(\tau), 0<\tau<t\} \tag{B-12}$$

is the same as (B-4) with F and G given by

$$F_t = \sum_{\xi_1'} \lambda\big(\xi_1, y(t) \,|\, \xi_1', y(t)\big)\pi_t(\xi_1') - \pi_t(\xi_1) \cdot \sum_{\xi_1'} \lambda\big(y(t) \,|\, \xi_1', y(t)\big)\pi_t(\xi_1') \tag{B-13}$$

$$G_t = \frac{\sum_{\xi_1} \lambda\big(\xi_1, y(t) \,|\, \xi_1', y(t-)\big)\pi_{t-}(\xi_1')}{\sum_{\xi_1'} \lambda\big(y(t) \,|\, \xi_1', y(t-)\big)\pi_{t-}(\xi_1')} - \pi_{t-}(\xi_1) \quad . \tag{B-14}$$

B.2 OUTLINE OF PROOF OF FILTERING EQUATIONS

A direct proof of these results depends on showing that (1) $\pi_t(\xi)$ has a derivative given by F_t at each time t at which J_t does not jump and that (2) at jump times t, $\pi_t(\xi)$ satisfies

$$\pi_t(\xi) = G_t + \pi_{t-}(\xi) \quad . \tag{B-15}$$

For jump times t, Eq. B-15 follows directly from Bayes rule for a finite-state Markov chain. We will sketch how to prove (1) in the rest of this section.

Assume that t>s and define $A_{s,t}$ as

$$\begin{aligned} A_{s,t} &= 1 \text{ if } J_t - J_s = 0 \\ &= 0 \text{ if } J_t - J_s > 1 \quad . \end{aligned} \tag{B-16}$$

Define ϕ_s as a function of sample paths of y as

$$\begin{aligned} \phi_s(y)(\tau) &= y(\tau) \text{ for } \tau \leq t \\ &= y(t) \text{ for } \tau > t \quad . \end{aligned} \tag{B-17}$$

Suppose that f is Y_t measurable $\left(Y_t = \sigma\{y(\tau), 0 \leq \tau \leq t\}\right)$. Then $f \cdot \phi_s$ is Y_s measurable and

$$f \cdot A_{s,t} = (f \cdot \phi_s) \cdot A_{s,t} \quad . \tag{B-18}$$

Therefore, letting $f = \pi_t(\xi)$, we get

$$\pi_t(\xi) \cdot A_{s,t} = \left(\pi_t(\xi) \cdot \phi_s\right) \cdot A_{s,t} \tag{B-19}$$

which implies that $\pi_t(\xi) \cdot A_{s,t}$ is Y_s V $\sigma\{A_{s,t}\}$ measurable. Thus,

$$\begin{aligned} \pi_t(\xi) \cdot A_{s,t} &= E\{\pi_t(\xi) | Y_s, A_{s,t}\} \cdot A_{s,t} \\ &= \Pr\{x(t)=\xi | Y_s, A_{s,t}\} \cdot A_{s,t} \quad . \end{aligned} \tag{B-20}$$

Note that

$$\Pr\{x(t)=\xi | Y_s, A_{s,t}=1\} = \frac{\Pr\{x(t)=\xi, A_{s,t}=1 | Y_s\}}{\Pr\{A_{s,t}=1 | Y_s\}} \tag{B-21}$$

where

$$\Pr\{x(t)=\xi, A_{s,t}=1 | Y_s\} = \sum_{\xi'} \Pr\{x(t)=\xi, A_{s,t}=1 | x(s)=\xi'\} \cdot \pi_s(\xi') \tag{B-22}$$

and

$$\Pr\{A^{s,t} | y_s\} = \sum_{\xi'} \Pr\{A_{s,t}=1 | x(s)=\xi'\} \cdot \pi_s(\xi') \quad . \tag{B-23}$$

Let $B_{s,t}$ denote the event that there is more than one jump of x in the interval (s,t], and let $B^c_{s,t}$ be the complementary event that there is, at most, one one jump in (s,t]. Then we have the relationships

$$\delta_{h(\xi)h(\xi')} \cdot \Pr\{x(t)=\xi, B^c_{s,t} \mid x(s)=\xi'\} \leq \Pr\{x(t)=\xi, A_{s,t}=1 \mid x(s)=\xi'\} \tag{B-24}$$

and

$$\delta_{h(\xi)h(\xi')} \cdot \Pr\{x(t)=\xi \mid x(s)=\xi'\} > \Pr\{x(t)=\xi, A_{s,t}=1 \mid x(s)=\xi'\} \tag{B-25}$$

It follows from Eq. B-24 and B-25 that

$$\left|\delta_{h(\xi)h(\xi')} \cdot \Pr\{x(t)=\xi \mid x(s)=\xi'\} - \Pr\{x(t)=\xi, A_{s,t}=1 \mid x(s)=\xi'\}\right| \leq \Pr\{B_{s,t} \mid x(s)=\xi'\} \quad . \tag{B-26}$$

Note that

$$\Pr\{B_{s,t} \mid x(s)=\xi'\} = o(t-s) \quad . \tag{B-27}$$

Putting Eqs. B-27 and B-26 together with Eq. B-3 gives

$$\Pr\{x(t)=\xi, A_{s,t}=1 \mid x(s)=\xi'\} = \delta_{h(\xi)h(\xi')} \cdot [\delta_{\xi\xi'} + (t-s) \cdot \lambda(\xi \mid \xi')] + o(t-s) \quad . \tag{B-28}$$

Using Eq. B-28, we can derive the approximations

$$\Pr\{x(t)=\xi, A_{s,t}=1 \mid Y_s\} = \delta_{h(\xi)y(s)} \cdot [\pi_s(\xi) + \sum_{h(\xi')=y(s)} \lambda(\xi \mid \xi') \cdot \pi_s(\xi') \cdot (t-s)] + o(t-s) \tag{B-29}$$

and

$$\Pr\{A_{s,t}=1 \mid Y_s\} = 1 + \sum_{h(\xi'')=h(\xi')=y(s)} \sum \lambda(\xi'' \mid \xi') \cdot \pi_s(\xi') \cdot (t-s) + o(t-s) \quad . \tag{B-30}$$

From Eqs. B-21, B-29, and B-30, we get

$$\Pr\{x(t)=\xi \mid Y_s, A_{s,t}=1\}$$

$$= \delta_{h(\xi)y(s)} \cdot \Big[\pi_s(\xi) + \sum_{h(\xi')=y(s)} \lambda(\xi|\xi')\cdot\pi_s(\xi')\cdot(t-s) \qquad \text{(B-31)}$$

$$- \sum_{h(\xi'')=h(\xi')=y(s)} \sum \lambda(\xi''|\xi')\cdot\pi_s(\xi')\cdot\pi_s(\xi)\cdot(t-s)\Big] + o(t-s) \quad .$$

Note that $\pi_s(\xi) = 0$ unless $h(\xi) = y(s)$. Thus,

$$\delta_{h(\xi)y(s)}\cdot\pi_s(\xi) = \pi_s(\xi) \qquad \text{(B-32)}$$

and Eq. B-21 can be expressed

$$\Pr\{x(t)=\xi \mid Y_s, A_{s,t}=1\}$$

$$= \pi_s(\xi) + \delta_{h(\xi)y(s)} \cdot \Big[\sum_{h(\xi')=y(s)} \lambda(\xi|\xi')\cdot\pi_s(\xi')\cdot(t-s) \qquad \text{(B-33)}$$

$$- \sum_{h(\xi'')=h(\xi')=y(s)} \sum \lambda(\xi'|\xi')\cdot\pi_s(\xi')\cdot\pi_s(\xi)\cdot(t-s)\Big] + o(t-s) \quad .$$

From Eq. B-20,

$$\pi_t(\xi) - \pi_s(\xi) = [\Pr\{x(t)=\xi \mid Y_s, A_{s,t}=1\} - \pi_s(\xi)]\cdot A_{s,t} + [\pi_t(\xi) - \pi_s(\xi)]\cdot(1 - A_{s,t}) \qquad \text{(B-34)}$$

and it follows from Eq. B-33 and the definition of F_s that

$$\pi_t(\xi) - \pi_s(\xi) = F_s\cdot(t-s) + o(t-s) + [\pi_t(\xi) - \pi_s(\xi)]\cdot(1 - A_{s,t}) \quad . \qquad \text{(B-35)}$$

Finally note that because y is right continuous,

$$\lim_{t\downarrow s} \frac{1-A_{s,t}}{t-s} = 0 \qquad \text{(B-36)}$$

is true for all s. If t is not a discontinuity of y, then it is also true that

$$\lim_{s \uparrow t} \frac{1-A_{s,t}}{t-s} = 0 \quad . \tag{B-37}$$

Note that π_t, π_s, and F_s are uniformly bounded. Using this fact together with Eqs. B-36 and B-37, it follows that Eq. B-34 implies π_t is continuous at t if t is not a discontinuity of y. Consequently, F_t is also continuous at such times t, so that

$$\lim_{s \to t} F_s = F_t \tag{B-38}$$

in Eq. B-34. It follows that π_t is continuously differentiable at t with derivative

$$\lim_{s \to t} \frac{\pi_t(\xi) - \pi_s(\xi)}{t-s} = F_t \quad . \tag{B-39}$$

B.3 EXAMPLE

As example, consider the four-state Markov process with state space {0,1}X{0,1} and transition rates given by (see Fig. B-1):

$$\begin{aligned} \lambda(0,1|0,0) &= \alpha, \quad \lambda(0,0|0,1) = \alpha' \\ \lambda(1,0|0,0) &= \beta, \quad \lambda(0,0|1,0) = \beta' \\ \lambda(1,1|1,0) &= \alpha, \quad \lambda(1,0|1,1) = \alpha' \\ \lambda(1,1|0,1) &= \gamma, \quad \lambda(0,1|1,1) = \gamma' \end{aligned} \tag{B-40}$$

and

$$\begin{aligned} \lambda(0,0|0,0) &= -(\alpha+\beta) \\ \lambda(1,0|1,0) &= -(\alpha+\beta') \\ \lambda(1,1|1,1) &= -(\alpha'+\gamma') \\ \lambda(0,1|0,1) &= -(\alpha'+\gamma) \quad . \end{aligned} \tag{B-41}$$

The $\lambda(\lambda|\xi_1', y')$ are given by

$$\begin{aligned} \lambda(1|0,0) &= \lambda(1|1,0) = \alpha \\ \lambda(0|0,0) &= \lambda(0|1,0) = -\alpha \\ \lambda(0|0,1) &= \lambda(0|1,1) = \alpha' \\ \lambda(1|0,1) &= \lambda(1|1,1) = -\alpha' \end{aligned} \tag{B-42}$$

In this example, $\pi_t(\xi_1)$ satisfies a linear differential equation with coefficients that switch randomly with $y(t)$. That is,

$$\dot{\pi}_t(0) = \beta' - (\beta + \beta')\pi_t(0) \text{ when } y(t) = 0 \tag{B-43}$$

and

$$\dot{\pi}_t(0) = \gamma' - (\gamma + \gamma')\pi_t(0) \text{ when } y(t) = 1 \quad . \tag{B-44}$$

Let θ_0 and θ_1 denote the equilibrium points of the individual differential equations (Eqs. B-43 and B-44), namely

$$\theta_0 = \frac{\beta'}{\beta + \beta'} \tag{B-45}$$

$$\theta_1 = \frac{\gamma'}{\gamma + \gamma'} \quad . \tag{B-46}$$

During an interval of time $(\tau_k, \tau_{k+1}]$ when $y(t) = i$, the probability $\pi_i(i)$ approaches θ_i exponentially. If $\theta_0 \neq \theta_1$, then $\pi_t(0)$ eventually becomes trapped between θ_0 and θ_1 and oscillates randomly as shown in Fig. B-2. If $\theta_0 = \theta_1$, then $\pi_t(0)$ approaches θ_i in the limit. If β, β', γ, γ' are very large compared to α, α', then the oscillations approach the limit

$$\theta_0 \cdot [1 - y(t)] + \theta_1 \cdot y(t) \quad . \tag{B-47}$$

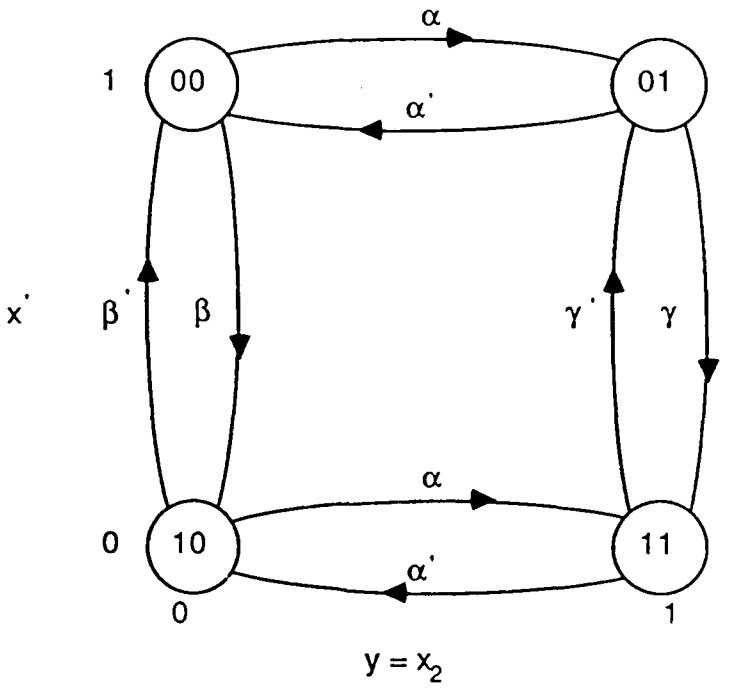


Figure B-1. State Transition Diagram

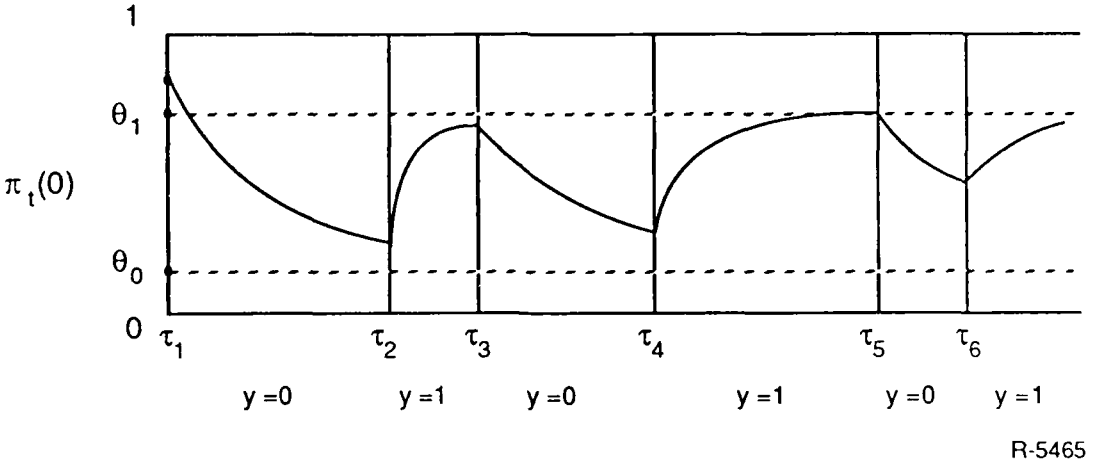


Figure B-2. Sample Path Behavior of $\pi_t(0)$

B.4 REGULAR MEASUREMENT MODEL

Suppose that $x_1(t)$ is itself a Markov process and $x_2(t)$ is a process that depends on $x_1(t)$ in such a way that $x(t) = (x_1(t), x_2(t))$ is a Markov process. Furthermore, suppose that for almost all sample paths of $x(t)$, $x_1(t)$ and $x_2(t)$ do not jump simultaneously. Then the transition rate λ for the joint process x has the following form

$$\begin{aligned} &\lambda(\xi_1,\xi_2|\xi_1',\xi_2') \\ &= \delta_{\xi_2\xi_2'}\cdot\lambda_1(\xi_1') \quad + \quad \delta_{\xi_1\xi_1'}\cdot\lambda_2(\xi_1',\xi_2') \end{aligned} \tag{B-48}$$

where λ_1 is the transition rate of x_1, and λ_2 is a transition rate for a process x_2 if ξ_1' is held constant, that is

$$\lambda_2(\xi_2|\xi_1',\xi_2') > 0 \text{ if } \xi_2 \neq \xi_2' \quad , \tag{B-49}$$

$$\sum_{\xi_2} \lambda(\xi_2|\xi_1',\xi_2') = 0 \tag{B-50}$$

for each ξ_1', ξ_2'. To see that Eq. B-48 is true note that if $x_1(t)$ and $x_2(t)$ do not have simultaneous jumps, then λ can be written in the form

$$\begin{aligned} &\lambda(\xi_1,\xi_2|\xi_1',\xi_2') \\ &= \delta_{\xi_2\xi_2'}\cdot a_1(\xi_1) \quad + \quad \delta_{\xi_1\xi_1'}\cdot a_2(\xi_2|\xi_1',\xi_2') \quad . \end{aligned} \tag{B-51}$$

Note that

$$\begin{aligned} 0 &= \sum_{\xi_1,\xi_2} \lambda(\xi_1,\xi_2|\xi_1',\xi_2') \\ &= \sum_{\xi_1} a_1(\xi_1|\xi_1',\xi_2') \quad + \quad \sum_{\xi_2} a_2(\xi_2|\xi_1',\xi_2') \quad . \end{aligned} \tag{B-52}$$

Define $\bar{a}_1$, $\bar{a}_2$ as

$$\bar{a}_1(\xi_1|\xi_1',\xi_2') = a_1(\xi_1|\xi_1',\xi_2') - \delta_{\xi_1,\xi_1'} \cdot \sum_{\xi_1} a_1(\xi_1|\xi_1',\xi_2') \quad , \tag{B-53}$$

$$\bar{a}_2(\xi_2|\xi_1',\xi_2') = a_2(\xi_2|\xi_1',\xi_2') - \delta_{\xi_2,\xi_2'} \cdot \sum_{\xi_1} a_2(\xi_2|\xi_1',\xi_2') \quad . \tag{B-54}$$

Then λ can be written as

$$\begin{aligned} &\lambda(\xi_1,\xi_2|\xi_1',\xi_2') \\ &= \delta_{\xi_2\xi_2'} \cdot \bar{a}_1(\xi_1|\xi_1',\xi_2') \quad + \quad \delta_{\xi_1\xi_1'} \cdot \bar{a}_2(\xi_2|\xi_1',\xi_2') \quad . \end{aligned} \tag{B-55}$$

where

$$\sum_{\xi_i} \bar{a}_i(\xi_i|\xi_1',\xi_2') = 0 \quad . \tag{B-56}$$

Summing Eq. B-55 over ξ_2 gives

$$\lambda(\xi_1|\xi_1') = \sum_{\xi_2} \lambda(\xi_1,\xi_2|\xi_1',\xi_2') = \bar{a}_1(\xi_1|\xi_1',\xi_2') \quad . \tag{B-57}$$

Substituting Eq. B-57 in Eq. B-55 gives the desired expression (Eq. B-48). Equation B-50 is just Eq. B-56, and Eq. B-49 follows from Eq. B-48. Note that Eq. B-48 can also be expressed by saying that the infinitesimal generator Λ of x given by

$$(\Lambda f)(\xi_1,\xi_2) = \sum_{\xi_1',\xi_2'} \lambda(\xi_1,\xi_2|\xi_1',\xi_2') \cdot f(\xi_1',\xi_2') \tag{B-58}$$

has the form

$$(\Lambda f)(\xi_1,\xi_2) = \Lambda_1 f(\cdot,\xi_2)(\xi_1) + \Lambda_2(\xi_1) f(\xi_1,\cdot)(\xi_2) \tag{B-59}$$

where Λ_1 is the infinitesimal generator corresponding to λ_1 and $\Lambda_2(\xi_1)$ is the infinitesimal generator corresponding to $\lambda_2(\xi_1,\cdot|\xi_1,\cdot)$.

We will call Eq. B-48 the regular measurement model. It is the discrete-state version of the following measurement model typically used for diffusion processes:

$$dx_1 = f(x_1)dt + dw \tag{B-60}$$

$$dx_2 = h(x_1,x_2)dt + dv \quad . \tag{B-61}$$

Note that the regular measurement model is only a special case of the partially-observed Markov process formulated in Eq. B-1. In some cases, it will be necessary to consider the more general model (e.g., the limit of a parameterized family of regular models need not be regular).

The finite-state filtering equation for the special measurement model (Eq. B-48) is given by

$$F_t = \sum_{\xi_1'} \lambda_1(\xi_1|\xi_1')\pi_t(\xi_1') + \lambda_2\big(y(t)|\xi_1,y(t)\big)\pi_t(\xi_1) - \pi_t(\xi_1) \cdot \sum_{\xi_1'} \lambda_2\big(y(t)|\xi_1',y(t)\big)\pi_t(\xi_1') \quad . \tag{B-62}$$

$$G_t = \frac{\lambda_2\big(y(t)|\xi_1,y(t-)\big)\pi_{t-}(\xi_1)}{\sum_{\xi_1'} \lambda_2\big(y(t)|\xi_1',y(t-)\big)\pi_{t-}(\xi_1')} - \pi_{t-}(\xi_1) \quad . \tag{B-63}$$

We are particularly interested in singular-perturbation problems where the observation process $y = x_2$ is speeded up by a factor of $\frac{1}{\varepsilon}$. This gives

$$\lambda(\xi_1,\xi_2|\xi_1',\xi_2') = \delta_{\xi_2\xi_2'}\cdot\lambda_1(\xi_1|\xi_1') + \frac{1}{\varepsilon}\cdot\delta_{\xi_1\xi_1'}\cdot\lambda_2(\xi_2|\xi_1',\xi_2') \quad . \tag{B-64}$$

The singularly perturbed filtering equations are

$$F_t = \sum_{\xi_1} \lambda_1(\xi_1)|\xi_1')\pi_t(\xi_1') + \frac{1}{\epsilon} \cdot \lambda_2(y(t)|\xi_1,y(t))\pi_t(\xi_1) - \frac{1}{\epsilon} \cdot \pi_t(\xi_1) \cdot \sum_{\xi_1'} \lambda_2(y(t)|\xi_1',y(t))\pi_t(\xi_1') \quad . \tag{B-65}$$

$$G_t = \frac{\lambda_2(y(t)|\xi_1,y(t-))\pi_{t-}(\xi_1)}{\sum_{\xi_1'} \lambda_2(y(t)|\xi_1',y(t-))\pi_{t-}(\xi_1')} - \pi_{t-}(\xi_1) \quad . \tag{B-66}$$

B.5 UNNORMALIZED FILTERING EQUATIONS

In nonlinear filtering for diffusion processes, Zakai's equation for an unnormalized conditional density is linear in the density and easier to analyze than the nonlinear equation for the conditional density itself. There is an analogous equation for an unnormalized conditional probability distribution in the finite-state continuous-time filtering problem. Consider the general problem formulated in Eq. B-1. Let $q_t(\xi)$ be the solution of the following stochastic differential equation.

$$dq_t(\xi) = [\delta_{h(\xi)y(t)} \cdot \sum_{\xi'} \lambda(\xi|\xi')q_t(\xi')]\cdot dt + [\delta_{h(\xi)y(t)} \cdot \sum_{\xi'} \lambda(\xi|\xi')q_{t-}(\xi') - q_{t-}(\xi)]\cdot dJ_t \quad . \tag{B-67}$$

Define $\bar{q}_t$ to be the normalization factor

$$\bar{q}_t = \sum_{\xi} q_t(\xi) \quad . \tag{B-68}$$

Then the conditional density $\pi_t(\xi)$ is

$$\pi_t(\xi) = \frac{q_t(\xi)}{\bar{q}_t} \quad . \tag{B-69}$$

It is not hard to show that Eq. B-69 is true for all t provided that it is true initially at t=0. At a jump time $t = \tau_k$

$$q_t(\xi) = \delta_{h(\xi)y(t)} \cdot \sum_{\xi'} \lambda(\xi|\xi')q_{t-}(\xi') \tag{B-70}$$

so that if Eq. B-69 is true for $t<\tau_k$, it is clearly true at τ_k (that is, Eq. B-70 preserves the relation (Eq. B-69) through the jump). In between jump times, $\tau_k \leq t < \tau_{k+1}$, $q_t(\xi)$ and $\bar{q}_t$ satisfy the ordinary differential equations

$$\dot{q}_t(\xi) = \delta_{h(\xi)y(t)} \cdot \sum_{\xi'} \lambda(\xi|\xi')q_t(\xi') \tag{B-71}$$

and

$$\dot{\bar{q}}_t = \sum_{\xi'} \lambda(y(t)|\xi')q_t(\xi') \quad . \tag{B-72}$$

It follows that the quotient $q_t(\xi)/\bar{q}_t$ satisfies the same equation as $\pi_t(\xi)$ in between jump times, namely

$$\begin{aligned} \dot{\pi}_t(\xi) = \delta_{h(\xi)y(t)} \cdot \Big[& \sum_{\xi'} \lambda(\xi|\xi')\pi_t(\xi') \\ & - \pi_t(\xi) \cdot \sum_{\xi'} \lambda(y(t)|\xi')\pi_t(\xi') \Big] \quad . \end{aligned} \tag{B-73}$$

Thus, if Eq. B-69 is true at $t = \tau_k$, it will remain true throughout the interval $\tau_k \leq t < \tau_{k+1}$.

For the measurement model given by Eq. B-11, the unnormalized distribution satisfies the equation

$$dq_t(\xi_1) = \Big[\sum_{\xi_1'} \lambda(\xi_1, y(t)|\xi_1', y(t))q_t(\xi_1')\Big]\cdot dt + \Big[\sum_{\xi_1'} \lambda(\xi_1, y(t)|\xi_1', y(t-))q_{t-}(\xi_1') - q_{t-}(\xi_1)\Big]\cdot dJ_t \quad . \tag{B-74}$$

For the regular measurement model, the unnormalized distribution satisfies the equation

$$dq_t(\xi_1) = \Big[\sum_{\xi_1'} \lambda_1(\xi_1|\xi_1')q_t(\xi_1') + \lambda_2(y(t)|\xi_1, y(t))q_t(\xi_1)\Big]\cdot dt + \Big[\lambda_2(y(t)|\xi_1, y(t-))q_{t-}(\xi_1) - q_{t-}(\xi_1)\Big]\cdot dJ_t \quad . \tag{B-75}$$

For the singularly-perturbed problem we have two choices: either to use the unnormalized equation (Eq. B-75) obtained by replacing λ_2 with $\frac{1}{\varepsilon}\lambda_2$ or to use a somewhat different version derived directly from Eqs. B-65 and B-66 in the same way as Eq. B-67 was derived. Note that there are many possible unnormalized distributions, and no one choice need be uniquely the best. We present the different version because it appears to give better scaling in the jump term.

$$\Big[\sum_{\xi_1'} \lambda_1(\xi_1|\xi_1')q_t(\xi_1') + \frac{1}{\varepsilon}\cdot\lambda_2(y(t)|\xi_1, y(t))q_t(\varepsilon_1)\Big]\cdot dt + \Big[\lambda_2(y(t)|\xi_1, y(t-))q_{t-}(\xi_1) - q_{t-}(\xi_1)\Big]\cdot dJ_t \quad . \tag{B-76}$$

Note that some care needs to be taken in analyzing the asymptotic behavior of π_t^ε by analyzing an unnormalized distribution such as q_t in Eq. B-76. If $\bar{q}_t$ has leading order magnitude $\psi(\varepsilon)$, then we need to analyze q_t to higher order, namely $o(\psi(\varepsilon))$, in order to say something useful about the asymptotics of π_t^ε.

APPENDIX C

DECISION PROBABILITIES FOR THE HYPOTHESIS TEST OF EQS. 6-57 AND 6-58

We are concerned here with analyzing the performance of the decision rule specified by Eqs. 6-57 and 6-58. For simplicity we drop the subscript k and we assume that x(0) = 1 (analogous expressions can be developed if x(0) = 0). Taking logarithms of Eqs. 6-57 and 6-58 and simplifying, we see that an equivalent form for this decision rule is for K even

$$C_1 R + \frac{1}{2} C_2 K \begin{array}{c} m_k = R \\ > \\ < \\ m_k = L \end{array} C_4 \qquad \text{(C-1a)}$$

while for K odd it is

$$C_1 R + \frac{1}{2} C_2 (K-1) + C_3 \begin{array}{c} m_k = R \\ > \\ < \\ m_k = L \end{array} C_4 \qquad \text{(C-1b)}$$

where

$$C_1 = (\beta-\alpha)\, g(\varepsilon) \quad , \qquad C_2 = \log\left[\left(1 + \frac{\alpha g(\varepsilon)}{\lambda_1}\right)\left(1 + \frac{\beta g(\varepsilon)}{\lambda_2}\right)\right]$$

$$C_3 = \log\left[1 + \frac{\alpha g(\varepsilon)}{\lambda_1}\right] \quad , \quad C_4 = \beta g(\varepsilon)\, T(\varepsilon) \quad . \qquad \text{(C-2)}$$

The key to the performance of Eq. C-1 is then the joint distribution for R and K under each hypothesis (i.e., $\rho(t)$ on the left or $\rho(t)$ on the right). For the case of $\rho(t)$ on the left

$$
\begin{aligned}
p(R,\ K=0) &= e^{-\lambda_1 T(\varepsilon)}\delta(R-T(\varepsilon)) \\
p(R,\ K=1) &= \lambda_1\, e^{(\lambda_2-\lambda_1)R}\, e^{-\lambda_2\, T(\varepsilon)} \\
p(R,\ K=2) &= (\lambda_1 R)\lambda_2\, e^{(\lambda_2-\lambda_1)R}\, e^{-\lambda_2\, T(\varepsilon)} \\
p(R,\ K=2M-1) &= \frac{(\lambda_1 R)^M}{M!}\ \frac{[\lambda_2(T(\varepsilon)-R)]^{(M-1)}}{(M-1)!}\ \lambda_2\, e^{(\lambda_2-\lambda_1)R}\, e^{-\lambda_2\, T(\varepsilon)} \\
p(R,\ K=2M) &= \frac{(\lambda_1 R)^M}{M!}\ \frac{[\lambda_2(T(\varepsilon)-R)]^{(M-1)}}{(M-1)!}\ \lambda_1\, e^{(\lambda_2-\lambda_1)R}\, e^{-\lambda_2\, T(\varepsilon)}
\end{aligned}
\qquad \text{(C-3)}
$$

where the last two expressions for for M = 2,3,... .

To illustrate the nature of the computations, suppose that $C_1,\ C_2 > 0$. Then

$$
\begin{aligned}
\phi_{LL}(\varepsilon) &= \sum_{M=0}^{\infty} \Pr\left(R < \frac{C_4-C_2M}{C_1},\ K = 2M\right) + \sum_{M=0}^{\infty} \Pr\left(R < \frac{C_4-C_3-C_2M}{C_1},\ K = 2M+1\right) \\
&= \sum_{M=0}^{N_1(\varepsilon)} \int_0^{T(\varepsilon)} p(R,2M)dR + \sum_{M=N_1(\varepsilon)+1}^{N_2(\varepsilon)} \int_0^{\gamma_1(\varepsilon)} p(R,2M)dR \\
&\quad + \sum_{M=0}^{N_3(\varepsilon)} \int_0^{T(\varepsilon)} p(R,2M+1)dR + \sum_{M=N_3(\varepsilon)+1}^{N_4(\varepsilon)} \int_0^{\gamma_2(\varepsilon)} p(R,2M+1)dR
\end{aligned}
\qquad \text{(C-4)}
$$

where

$$N_1(\varepsilon) = \frac{C_4 - C_1\ T(\varepsilon)}{C_2} \quad , \quad N_2(\varepsilon) = \frac{C_4}{C_2}$$

$$N_3(\varepsilon) = \left[\frac{C_4 - C_3 - C_1\ T(\varepsilon)}{C_2}\right] \quad , \quad N_4(\varepsilon) = \left[\frac{C_4 - C_3}{C_2}\right] \qquad \text{(C-5)}$$

$$\gamma_1(\varepsilon) = \frac{C_4 - C_2 M}{C_1} \quad , \quad \gamma_2(\varepsilon) = \frac{C_4 - C_3 - C_2 M}{C_1} \quad .$$

END

DATE

FILMED

3 — 88

DTIC